ExceleRate Illinois

Validation Study Report

December 2017

Frank Porter Graham Child Development Institute

University of North Carolina at Chapel Hill

ExceleRate Illinois: Validation Study Report

Noreen Yazejian, Ph.D. (Principal Investigator)
Margaret R. Burchinal, Ph.D. (Co-PI)
& Sandra Soliday Hong (Co-PI)

Frank Porter Graham Child Development Institute
The University of North Carolina at Chapel Hill

Eboni Howard, Ph.D. (Subcontract PI)

American Institutes for Research

Final Report Submitted to
Illinois State Board of Education
December 31, 2017

Acknowledgments:

We are grateful to the Illinois early care and education providers, teachers, families, and children who participated in this study. We thank them for their perspectives and their work on behalf of Illinois' children and families.

Validation Study team:

FPG
Noreen Yazejian, Ph.D.
Margaret Burchinal, Ph.D.
Sandra Soliday Hong, Ph.D.
Ciani Bush-Johnson
Laura Kuhn, Ph.D.
Sarah Wackerhagen
Eric Savage

AIR
Eboni Howard, Ph.D.
Patricia Garcia-Arena, Ph.D.
Lisa Sturgis
Emily Anderson

Suggested Citation:
Yazejian, N., Burchinal, M.R., Soliday Hong, S., & Howard, E. (2017). ExceleRate Illinois: Validation Study Report. Chapel Hill, NC: Frank Porter Graham Child Development Institute, University of North Carolina.

Table of Contents

Executive Summary 1

Section 1. Introduction 3

1.1. Background on QRIS and QRIS Validation 3

1.2. Background on ExceleRate Illinois 5

1.3 Conceptual Framework 6

1.4 Research Questions: Validation of ExceleRate Illinois 6

Section 2. Methods 7

2.1. Sample and Data Collection 7

2.2. Measures 14

2.3. Data Analysis 19

Section 3. Results 21

3.1. Descriptive Statistics for FPG/AIR-Gathered Data 21

3.2. Descriptive Statistics for Existing ExceleRate Data 27

3.3. Research Question 1: To What extent does ExceleRate Illinois differentiate levels of quality in programs overall and across each domain? 30

3.4. Research Question 2: To what extent does a program's score on each domain contribute independently to its overall rating? 36

3.5. Research Question 3: What combination and/or weighting of indicators best discriminate levels of quality? 37

3.6. Research Question 4. Do children who participate in programs rated at higher Circles of Quality overall and across particular domains demonstrate greater gains in child outcomes compared to children who participate in programs rated at lower levels overall and across domains? 40

Section 4. Summary and Conclusions 44

4.1. Key Findings 44

4.2. Study Limitations 46

4.3. Conclusions and Recommendations for Consideration 48

Section 5. References 50

List of Tables

Table 2-1. Sample of Schools and Centers by Region, Pathway, and Circle of Quality 8
Table 2-2. Director and Center/School-Level Characteristics 9
Table 2-3. School and Center Curricula 10
Table 2-4. Number of Child Assessments and Classroom Observations 10
Table 2-5. Teacher characteristics 11
Table 2-6. Classroom Staffing and Enrollment by Age Group 12
Table 2-7. Number of Programs, Classrooms, and Children with Child Assessment Data ... 12
Table 2-8. Child and Family Characteristics 13
Table 2-9. Alignment between ExceleRate Illinois Standards and Criteria and Study Measures 14
Table 3.1 Number of Classrooms in Validation Study 21
Table 3-2. Classroom Quality Scores at the Center/School Level 22
Table 3-3. Child Outcome Scores by Time Point and Circle of Quality 23
Table 3-4. Family Experiences with the Center/School 24
Table 3-5. Director Report of Quality Improvement Activities 25
Table 3-6. Director Report of Inclusion of Children with Special Needs and Dual Language Learners 26
Table 3-7. Director Report of Assessments and Communication with Families 27
Table 3-8. Descriptive Statistics for ExceleRate Domain Scores 28
Table 3-9. Correlations among Standards Met within Domains and ECERS-R Scores (n=327-433 centers/schools) 29
Table 3-10. Comparison of Quality Scores by Circles of Quality 30
Table 3-11. Average Program-Level Observed Quality by ExceleRate Circle of Quality 32
Table 3-12. Differences in Mean Classroom Quality Scores across Three Circles of Quality, Region, and Program Type 32
Table 3-13. Differences in Mean Program Quality Scores across Three Circles of Quality, Region, and Program Type 33
Table 3-14. Differences in Mean Classroom Quality Scores across Three Circles of Quality, Pathway, and Program Type 35
Table 3-15. Differences in Mean Program Quality Scores across Three Circles of Quality, Pathway, and Program Type 35

Table 3-16. Associations between ExceleRate Domain Scores and the Likelihood of Receiving a Gold Circle of Quality Rating 37
Table 3-17. Factor Loadings for ISBE Checklist Items by Quality Domain 38
Table 3-18. Correlations between Classroom and Program Quality Measures and ExceleRate Rating Checklist Standards, Domains, and Quality Assessments 39
Table 3-19. Correlations between changes in Child Outcome Scores and ExceleRate Rating Checklist Standards, Domains, and Quality Observations 40
Table 3-20. Child Outcomes, Fall and Spring, by ExceleRate Circle of Quality 41
Table 3-21. HLM Analyses of Child Outcomes by ExceleRate Circles of Quality 42

List of Figures

Figure 3-1. Average ExceleRate Domain Scores and Total Standard Scores by Circle of Quality 31

Figure 3-2. Comparison of Adjusted Means across Classroom and Program Quality Measures by Circle of Quality 34

Executive Summary

The ExceleRate Illinois tiered quality rating and improvement system was examined in a validation study conducted by the Frank Porter Graham Child Development Institute (FPG) in collaboration with American Institutes for Research (AIR). ExceleRate Illinois provides systems for monitoring and improving child care quality and providing clear information to families about the quality of the early care and education (ECE) of enrolled centers and schools. This study sought to examine the validity and reliability of the ExceleRate Circles of Quality. The validation study answered the following questions:

- To what extent does ExceleRate Illinois differentiate levels of quality in programs overall and across each domain?
- To what extent does a program's score on each domain contribute independently to its overall rating?
- What combination and/or weighting of indicators best discriminate levels of quality?
- Do children who participate in programs rated at higher Circles of Quality overall and across particular domains demonstrate greater gains in child outcomes compared to children who participate in programs rated at lower levels overall and across domains?

These issues were addressed using two sources of data: existing ExceleRate ratings of centers and schools and data gathered by the FPG/AIR team on a sample of participating programs. The existing ExceleRate data included ratings on the standards[1] and a widely used quality rating scale (the Early Childhood Environment Rating Scale-Revised). Data were gathered by the FPG/AIR team from 153 programs randomly selected to represent the ExceleRate Circles of Quality, regions of the state, type of programs (Preschool for All or other center-based programs), and pathway (assessed with the revised rating system or accredited based on meeting other requirements). Classroom and program quality data were collected, and child assessments were conducted using English or Spanish standardized measures.

Results indicated that ExceleRate successfully differentiated program quality. Analyses of ExceleRate standards indicated that programs rated at the Gold Circle of Quality scored substantially higher on all four ExceleRate quality domains and on observed classroom quality. The data gathered independently by the FPG/AIR team found that programs in higher Circles of Quality were rated as showing higher levels of preschool instructional support, infant classroom global quality, and center-level support of child assessments, family partnership, and program management.

In analysis of ExceleRate data on Preschool for All (PFA) programs, two of the four quality domains, Teaching and Learning and Leadership and Management, contributed independently to predicting a Gold Circle of Quality rating. The other two domains, Family and Community

[1] Standards data were available on Preschool for All programs, but not on other center-based programs.

Engagement and Qualifications and Continuing Education, did not; however, because these analyses were completed using data on PFA programs only, there was limited variability in the Qualifications and Continuing Education domain. All PFA teachers are required to have Bachelor degrees, and therefore the study was limited in its ability to detect associations among that domain and overall ratings. It appears that four scales, not a single summary score, may be needed to best summarize and communicate the ExceleRate standards. Analyses indicated four separate dimensions, labeled as classroom quality, structural quality, management, and quality improvement. Finally, no evidence indicated that children showed larger gains from fall to spring (~5 months) in cognitive or social outcomes measured in the study when they attended programs with higher Circles of Quality. The limitations of the study (see Section 4.2) should be taken into account when considering the study findings.

These findings provide strong validation of ExceleRate Illinois in terms of reliably measuring the quality of ECE centers and schools, using measures that are widely agreed to be important for children and families. Further study would be needed to understand how selected child learning outcomes are related to ExceleRate measures and activities. As with the other state rating systems, analyses of ExceleRate suggested that early childhood quality is not unidimensional. A single rating score makes the system easier for parents to understand and use, but may obscure associations with child outcomes if only some of the dimensions are actually related to the selected outcomes.

Section 1. Introduction

ExceleRate Illinois, a quality rating and improvement system (QRIS), was conceptualized and designed starting in 2012 through support from a federally funded Race to the Top – Early Learning Challenge (RTT-ELC) grant to implement an integrated system of high-quality early learning programs and services. The intent of the RTT-ELC grant was to increase enrollment of low-income and disadvantaged children in high quality early care and education (ECE) programs (more information about the broader RTT-ELC activities are available in Schilder, 2017). Efforts to strengthen the ECE system in Illinois were coordinated by the Illinois Governor's Office of Early Childhood Development (OECD) in partnership with the Illinois State Board of Education (ISBE), the Illinois Department of Human Services (DHS), and the Illinois Department of Children and Family Services (DCFS).

A validation study of ExceleRate Illinois was designed to examine whether the tiers of ExceleRate meaningfully differentiate levels of ECE quality and whether those levels represent meaningful differences in children's experiences that are reflected in increased school readiness and social development for children enrolled in ExceleRate schools and centers. The Frank Porter Graham Child Development Institute (FPG) at the University of North Carolina at Chapel Hill in collaboration with the American Institutes for Research (AIR) completed the validation study of ExceleRate Illinois at the request of ISBE during the 2015-16 school year. This final report provides a summary of the activities and results from the validation study. This first section provides an introduction to the study, including background information on QRISs, a description of ExceleRate Illinois, and the study's conceptual framework and research questions. The remainder of the report is organized as follows: Section 2. Methods, Section 3. Results, and Section 4. Discussion of Results.

1.1. Background on QRIS and QRIS Validation

Motivated in part by ECE research identifying quality indicators associated with children's outcomes, states have implemented QRISs to enhance ECE program quality. Nearly all state QRISs include staff training and education and the classroom or learning environment (although the latter is only measured at higher levels of quality in some states). States differ on whether and to what extent they include parent engagement activities, business practices, child-staff ratios, or national accreditation status. QRISs serve multiple purposes, one of which is to provide a standard way of rating program quality, based on multiple criteria, and making the rating information available to parents. The assumption underlying this function of QRISs is that parents often lack good information about program quality and if such information was available they would be more likely to choose higher-rated settings. As a result, lower-quality providers would be given an incentive to either improve the quality of their program or to leave the market (Zellman & Perlman, 2008). QRISs also represent a systematic approach to providing a range of technical assistance, resources, and incentives for programs to improve program quality. Such efforts include consultation around quality improvement, increased investments for professional development scholarships, micro grants for other targeted quality improvement efforts, and in some instances higher levels of subsidy payments for more highly rated programs. The goal of these efforts is to foster and support providers' efforts to improve

the quality of care they provide. Thus, QRISs attempt to improve quality by affecting both the demand for high-quality care and the supply of such care. Of course, the success of such efforts rests on the ability of rating systems to accurately identify and measure key aspects of quality and the willingness of providers to participate in a rating system (Zellman & Perlman, 2008).

Results from recent QRIS validation studies suggest that rating systems are succeeding in discriminating higher quality from lower quality ECE programs based on observational measures of classroom quality (Tout et al., 2016; Yazejian et al., forthcoming). Studies that involve child outcomes so far provide mixed results, which is not surprising given the diversity of rating systems and related policies across states. A small study in Missouri found that low-income children in higher rated programs learned more than their peers in lower rated programs (Thornburg, Mayfield, Hawks, & Fuger, 2009). However, a large study of Colorado's rating system did not find that children's gains in school readiness differed systematically as a function of star rating level (Zellman, Perlman, Le, & Setodji, 2008). More recently, Karoly's literature review of 12 QRIS evaluations found that programs with higher ratings had higher Environment Rating Scale (ERS) scores (Karoly, 2014). Seven of the evaluations included a measure of quality that was not included as part of the rating; all of these found small but statistically significant associations between quality and ratings. Four of the twelve studies identified in the literature review had strong research designs for examining relations between ratings and children's skills, and two of these studies found linkages between QRIS ratings and child outcomes.

A secondary analysis of a large study of quality and child outcomes applied the rating criteria from several states and asked whether children showed larger gains when their classroom met higher standards. They reported associations between individual quality variables and both observed quality and child outcomes, but few associations were observed between overall star ratings and either observed quality or child outcomes (Sabol, Hong, Pianta, & Burchinal, 2013). Using a similar approach, a meta-analysis of Head Start, pre-kindergarten, and nationally representative data reported statistically significant, but very small, associations among a simplified star rating system and child outcomes (Burchinal, Hong, Sabol, & Forestieri, 2014). Concerned that quality is multidimensional, Burchinal and colleagues selected classroom quality indicators with empirical evidence relating them to child outcomes and created a point rating system. These results may provide further evidence that continuous improvement of QRISs based on evaluations are likely to improve their ability to predict child outcomes.

Early care and education researchers define QRIS validation as a multi-step process that includes multiple studies, analyses, and sources of information and that does not necessarily lead to a yes/no designation of validity (Zellman & Fiene, 2012). The current study provides one piece of evidence for stakeholders in Illinois to consider as they continue to build the state's early childhood system.

1.2. Background on ExceleRate Illinois

ExceleRate Illinois is a statewide QRIS for early childhood programs that grew out of an earlier child care rating system. Implementation of the new system began in 2013. The focus of ExceleRate Illinois is to provide the following:

- A consistent definition of what constitutes high-quality early learning and development programs across a diverse range of settings,
- An accountability framework for programs receiving public funding,
- Pathways and support for continuous improvement, and
- Useful information for parents and families to help them select a high-quality program that best meets their needs.

ExceleRate Illinois uses a block rating structure and is composed of four levels of sequentially increasing quality: Licensed, Bronze, Silver, and Gold Circles of Quality. Licensed center-based programs and family child care homes are automatically enrolled once licensed. Participation at higher levels is voluntary. As a block system, schools and centers must meet all of the requirements of one level before moving up to the next level. Programs are rated based on several specific standards, or indicators, in four domains: (1) Teaching and Learning (6 standards covering learning environment, curriculum, instructional quality, child screening, child assessment, and inclusion of children with special needs); (2) Family and Community Engagement (2 standards covering family and community engagement and transitions), (3) Leadership and Management (4 standards covering program administration, ratios and group sizes, continuous quality improvement, and culturally and linguistically appropriate practice), and (4) Qualifications and Continuing Education (3 standards covering director qualifications, teacher qualifications, and staff development). More information on the standards at each Circle of Quality can be found at http://www.excelerateillinoisproviders.com/docman/resources/13-overview-of-charts/file.

Programs have options, or pathways, for how they can meet standards in each of these four domains, and these options vary depending on program type. Head Start programs can meet standards at the Silver or Gold Circle of Quality depending on their most recent Head Start monitoring results which include scores on the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008). Preschool for All (PFA), the state's Pre-Kindergarten program, and Prevention Initiative programs can also meet standards at the Silver or Gold Circle of Quality depending on their most recent PFA monitoring results, including their scores on the Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford, & Cryer, 1998). Any licensed, Head Start, or PFA program also can apply for the Silver or Gold Circle if they are accredited by specific national accrediting bodies (e.g., the National Association for the Education of Young Children or National Accreditation Commission) using current accreditation status and supplemental documentation as required. These options are all considered an "accredited" pathway because Head Start, PFA, or other accrediting body standards, with documentation, are used to determine ratings. The "assessed" pathway involves an on-site assessment by a state-approved assessor to determine ratings. Thus, ExceleRate Illinois is a consistent system that also provides flexibility to encourage maximum participation.

1.3 Conceptual Framework

Three key promises of a QRIS are: (1) to ensure that parents and other stakeholders can select the highest quality child care programs for children based on meaningful quality ratings; (2) to promote quality improvement in all child care settings through the provision of important benchmarks of quality measured periodically to examine change over time; and (3) to provide quality that can support children's optimal development (Zellman & Fiene, 2012). In order to make good on these promises, a QRIS must be validated. Validation is a multi-faceted process that involves demonstrating a high degree of correlation between ratings and important indicators of program quality (construct validity), a high degree of correlation between ratings and important measures of quality not included as indicators in the rating system (convergent validity), a high degree of correlation between quality ratings and desired child outcomes (predictive validity), and a significant capacity of the ratings system to distinguish high- and low-quality sites (discriminant validity). In addition, the validation process assumes reliability in quality measures, an assumption that requires indicators and quality ratings to not be biased or flawed in their reflection of quality status. Given these definitions, we conducted a multi-pronged validation of ExceleRate Illinois to answer the validation study questions specified by the state of Illinois.

1.4 Research Questions: Validation of ExceleRate Illinois

1. **To what extent does ExceleRate Illinois differentiate levels of quality in programs overall and across each domain?**
   - 1.1. To what extent do programs at higher tiers in ExceleRate Illinois demonstrate higher levels of quality overall and within each domain?
   - 1.2. According to independent measures of process quality, to what extent do the programs at the highest Circle of Quality demonstrate higher quality compared to other participating programs? Do the programs with Gold ratings have higher quality according to independent measures than other participating programs?
   - 1.3. Do associations among Circle of Quality rating and quality as measured independently vary by program type (e.g., licensed child care setting, Head Start, Preschool for All) or "pathway" to Circle of Quality level?

2. **To what extent does a program's score on each domain contribute independently to its overall rating?**
   - 2.1. Which quality domains (e.g., Teaching & Learning, Family & Community Engagement) and subdomains (e.g., Learning Environment, Curriculum) best predict ExceleRate Circle of Quality rating? Is the number of standards met within a given domain higher in programs at higher Circles of Quality for each domain?

3. **What combination and/or weighting of indicators best discriminate levels of quality?**
   - 3.1. To what extent does each of the indicators contribute to the overall domain rating? Are there indicators that appear to be more strongly related to the overall domain score computed without that indicator or to independent measures of process quality?

3.2. Is there evidence that the indicators might be multidimensional, suggesting that combining them with other indicators within that domain could be problematic in terms of a continuous quality improvement model?
3.3. Do we see evidence that the domains might be multidimensional, suggesting the need for more than one total score (e.g., Family & Community Engagement might be important, but may not be related to classroom structural and process quality)?
3.4. Does weighting the domains and indicators produce total scores that are more strongly related to independent quality measures and child outcomes?

4. Do children who participate in programs rated at higher Circles of Quality overall and across particular domains demonstrate greater gains in child outcomes compared to children who participate in programs rated at lower levels overall and across domains?

4.1. Adjusting for family characteristics, to what extent are the gains in children's academic and social skills larger among children who attended programs rated at higher Circle of Quality levels compared to children in programs rated at lower levels?

4.1a To what extent does each domain predict gains in child outcomes?

4.1b To what extent does each standard predict gains in child outcomes?

4.2. When children attend programs rated at higher Circles of Quality, are gains larger for children from low-income families, ethnic minority children (e.g., African American, Latino), from families in which English is a second language, or who have special needs than for other children?

4.3. Do child gains related to attending programs rated at higher Circles of Quality vary depending on type of program (e.g., licensed child care setting, Head Start, Preschool for All) or "pathway" to Circle of Quality level?

Section 2. Methods

Two sources of data were used to answer the research questions: existing data and data gathered by the FPG/AIR team. Existing data consisted of ExceleRate ratings of programs enrolled in IL's prekindergarten program, Preschool for All (PFA), shared with the research team by the Illinois State Board of Education (ISBE), and ExceleRate ratings of other center-based programs, shared with the research team by the Illinois Network of Child Care Resource and Referral Agencies (INCCRRA). This section of the report presents detailed information the data gathered by the FPG/AIR team, including sampling, data collection procedures, and data collection measures.

2.1. Sample and Data Collection

We selected a stratified random sample of 175 centers representing each of the four Circles of Quality, region of the state, and pathway (for Silver and Gold programs). We randomly selected infant/toddler and preschool classrooms in a similar proportion as does ExceleRate in the QRIS assessments to conduct classroom observations, teacher surveys, and director interviews to describe the quality of ExceleRate programs, selecting about one-third of the classrooms within a center with a maximum of 5 classrooms per center. We also collected data from 4-6 preschool children per classroom in the fall and spring to describe change in child outcomes among

children attending ExceleRate programs. The child assessments included direct assessments of language, academic, and self-regulation skills; teacher ratings of social and academic skills; and parent ratings of health. We used standardized measures, with selected measures having a Spanish version.

School and Center characteristics. Sampling involved four stratification factors: Type (PFA or other center programs), Region (Chicago, Collar Counties, Downstate urban, Downstate rural), Circle of Quality (Gold, Silver, Licensed), and Pathway (assessed or accredited). 175 schools and centers initially agreed to participate in the study at the beginning of the school year, and during the uncertainty of the state budget in the fall of 2015, 22 programs dropped from the study when contacted to schedule data collection. Despite these challenges, we were able to recruit sufficient numbers of programs for each stratification factor (see Table 2-1).

- Type of program: 102 Centers, 51 PFA programs;
- Region: 33 Chicago, 37 Collar Counties, 64 Downstate Urban, 19 Downstate Rural;
- Circle of Quality: 24 Licensed, 50 Silver, 79 Gold; and
- Pathway: 41 Accredited, 112 Assessed.

Table 2-1. Sample of Schools and Centers by Region, Pathway, and Circle of Quality

		Accredited Circle of Quality			Assessed Circle of Quality				
Region	**Program Type**	**Silver**	**Gold**	**Total**	**Licensing**	**Silver**	**Gold**	**Total**	**Total**
Chicago	**Center**	0	12	12	5	4	0	**9**	**21**
	School	0	0	**0**	0	6	7	**13**	**13**
	Total	0	12	12	5	10	7	**22**	**32**
Collar Counties	**Center**	1	8	**9**	8	10	4	**22**	**31**
	School	0	0	**0**	0	2	5	**7**	**7**
	Total	1	8	**9**	8	12	9	**39**	**48**
Downstate Urban	**Center**	1	17	**18**	8	10	5	**23**	**41**
	School	0	0	**0**	0	8	14	**22**	**22**
	Total	1	17	**18**	8	18	19	**45**	**63**
Downstate Rural	**Center**	0	2	**2**	3	3	0	**6**	**8**
	School	0	0	**0**	0	4	6	**10**	**10**
	Total	0	2	**2**	3	7	6	**16**	**18**
Total	**Center**	2	39	**41**	24	27	9	**60**	**101**
	School	0	0	**0**	0	20	32	**52**	**52**
	Total	2	39	**41**	**24**	**47**	**41**	**112**	**153**

Twenty-eight percent of centers and schools reported being accredited by the National Association for the Education of Young Children (NAEYC) and 8% by the National Accreditation Commission (NAC). The final total sample was 153 programs, and the overall acceptance rate, calculated as the number that accepted divided by the number that were sampled and contacted (1,201), was 13%. Anecdotal reports from program directors suggest that the state budget uncertainty occurring in 2015 when we were recruiting may have contributed to this low acceptance rate; the low rate of acceptance should be taken into account when considering the extent to which the study findings apply to all programs in the state.

Director characteristics. A total of 151 ECE directors participated in interviews about their program with demographic questions about themselves. The majority of the directors were White, had extensive experience teaching young children, and had a bachelor's degree or higher, mostly in a non-education field (see Table 2-2).

Table 2-2. Director and Center/School-Level Characteristics

	n	M/%
Total Participants		
Female	151	93%
Race (check all that apply)		
American Indian or Alaska Native	151	0%
Asian/ Asian American	151	3%
Black/ African American	151	19%
Pacific Islander	151	1%
White	151	79%
Other	151	2%
Ethnicity		
Hispanic/ Latino	151	4%
Education		
High School Diploma/GED	151	2%
Associate Degree	151	11%
BA/BS Degree	151	32%
MA/MS Degree	151	55%
Proportion of Teachers with a Degree in the Center		
Early childhood Education	151	26%
Elementary Education	151	17%
Special Education	151	6%
Clinical/counseling Psychology	151	2%
Early childhood Special Education	151	1%
Another Field of Education	151	15%
Another Field	151	48%
Received ISBE Coaching	128	57%

	n	M/%	
Proportion of Teachers in Center Who Speak a Language Other than English at Home	151	9%	
Annual Salary	151	$53,027	$18,269
ECE Experience (Years)	151	19.38	9.63
Full-Time	151	97%	

Almost all of the centers and schools reported using a curriculum, primarily Creative Curriculum or other curricula not listed—commonly locally developed curricula (see Table 2-3).

Table 2-3. School and Center Curricula

	n	%
Research Based Curriculum used	151	92%
Staff Training Provided	139	97%
Curriculum		
Creative Curriculum	151	73%
Other	151	30%
High Scope	151	3%
Tools of the Mind	151	1%
OWL	151	0%
Bank Street	151	0%
We Can	151	0%

Teacher characteristics. Classrooms were randomly selected within programs, and stratified so that half the selected sample was comprised of infant-toddler classrooms and half preschool classrooms if both types were present. The goal was to select one-third of the number of classrooms within a program, with a maximum of 5 classrooms per program. A total of 357 classrooms were included in the study (225 preschool and 132 infant/toddler). Observations were conducted in 357 infant, toddler, preschool/Pre-K, and mixed-age classrooms, and child assessments were conducted with preschool-aged children in 219 classrooms (see Table 2-4).

Table 2-4. Number of Child Assessments and Classroom Observations

Age Group	Number of Classrooms with Observations	Number of Classrooms with Child Assessments
Infant	56	0
Toddler	63	0
Preschool/ Pre-K	219	213
Infants/Toddlers: Mixed	13	0
Preschool/ Pre-K: Mixed	6	6
Total	**357**	**219**

Teachers were asked to complete a survey to gather information about their characteristics. The survey included teacher demographics and teacher qualifications as well as information about their classroom, such as curricula used, group size/ratio, and number of children with disabilities. We also gathered information about family engagement practices. Table 2-5 summarizes the demographic information about teachers. Teachers were mostly female, White, experienced, full-time teachers, with high levels of education.

Table 2-5. Teacher characteristics

	n	%/Mean	Mean
Total Participants			
Female	341	99%	
Race (check all that apply)			
American Indian or Alaska Native	341	2%	
Asian/ Asian American	341	11%	
Black/ African American	341	32%	
Pacific Islander	341	1%	
White	341	82%	
Other	341	16%	
Ethnicity			
Hispanic/ Latino	341	13%	
Education			
High School Diploma/GED	340	14%	
Associate Degree	340	21%	
BA/BS Degree	340	45%	
MA/MS Degree	340	21%	
Teacher College Degree			
Early childhood Education	292	64%	
Elementary Education	292	20%	
Special Education	292	6%	
Clinical/counseling Psychology	292	2%	
Early childhood Special Education	292	4%	
Another Field of Education	292	9%	
Another Field	292	20%	
Received ISBE Coaching	342	24%	
Teacher speaks language other than English at home	341	14%	
Annual Salary	331	$34,789	$13,221
ECE Experience (Years)	332	11.81	8.68
Full-Time	332	89%	

The average reported class sizes were 8 for infants, 10 for toddlers, 12 for two-year-olds, and 17 for preschool/Pre-K. Classrooms across the age ranges had two or three teachers on average (see Table 2-6).

Table 2-6. Classroom Staffing and Enrollment by Age Group

Ages Served	# of children	# Lead Teacher (s)	# Assistant Teacher(s)	# Other Adult (s)	Total Teachers/ Adults
Infants (0-14 mos)	8.06	1.42	1.26	0.11	2.79
Toddlers (15-23 mos)	9.67	1.33	1.06	0.13	2.52
Twos (25-35 mos)	12.39	1.29	0.95	0.09	2.33
Preschool (3 yrs -K)	17.49	1.23	0.98	0.21	2.42

Child Pre-Academic and Social Skills. Teachers in the recruited preschool classrooms were sent packages of parent consent forms. Programs serving Spanish-speaking families were sent copies of the consent forms in Spanish as well as in English. Teachers were asked to send home the consent forms and retain the returned forms. The data collectors obtained the returned consent forms from the teacher, and data collection was scheduled. The data collector examined the parent consents, eliminating children not eligible due to age (older or younger than 3-5 years of age) or language (child spoke neither English nor Spanish) or who were not present that day. Data collectors then sorted the consents based on gender and home language and used a random number list to identify which children in each of the possible four groups to assess. In cases where a child assessed in the fall could not be assessed in the spring, another child with parent permission was assessed when possible. This resulted in a sample of 731 children in the fall, 718 in the spring, and a total sample size of 877 preschool children (see Table 2-7). Five percent of the sample was assessed in Spanish (n=44), and of those about two-thirds had assessments in both the fall and spring (n=28). The average length of time between fall and spring assessments was 4.96 months (*SD*=1.12 months).

Table 2-7. Number of Programs, Classrooms, and Children with Child Assessment Data

Completed assessments	Programs	Classrooms	Children
Fall only assessments	6	12	159
Spring only assessments	5	12	146
Both Fall and Spring assessments	142	195	572
Total	153	219	877

The protocol involved direct assessment of children's language, academic, and self-regulation skills; teacher ratings of academic and social skills; and parent ratings of child health. The data collectors administered a screening protocol, the Pre-LAS® (Duncan & DeAvilla, 1998), to children whose parents or teachers reported that a language other than English was spoken at

home. Children who passed the Pre-LAS, along with all children whose parents spoke only English at home, were assessed in English. If children did not pass the Pre-LAS, they were assessed in Spanish if parents or teachers reported Spanish was spoken at home. Assessments were collected on laptops, and the teacher was sent an email with a link to a website with questionnaires for her to complete about herself and each study child. Teachers without children in the study (i.e., infant/toddler teachers) were sent a link to the website to complete the questionnaire about themselves. Paper surveys were also sent to teachers who did not respond electronically. A questionnaire and self-addressed stamped envelope was sent home for the parents to complete and return. Sample demographics (see Table 2-8) generally reflected the population of children under 5 in Illinois (Illinois Early Childhood Asset Map, http://iecam.illinois.edu/). Note that about a third of the sample did not provide information on income and therefore the average may not be representative.

Table 2-8. Child and Family Characteristics

	n	%	Mean	SD
Male	751	48%		
Race (check all that apply)				
American Indian/Alaska Native	729	1%		
Asian American/ Pacific Islander	729	6%		
Black/African American	729	21%		
White	729	68%		
Other	729	2%		
Ethnicity				
Hispanic	739	23%		
Mother/primary caregiver highest education level				
Some High School	721	7%		
High School diploma	721	17%		
Some College	721	24%		
Associate Degree	721	11%		
BA/BS Degree	721	22%		
MA/MS Degree	721	16%		
PhD Degree	721	3%		
Household Income	581		$62,892	$41,491
Child Health	467		4.54	.65
Child has IEP/IFSP	730	7%		
Child receives speech and language services	465	16%		
Dual Language Learners	744	30%		
Age child enrolled in center/school	447		2.68	1.11
Hours per week child attends center/school	458		25.57	14.69

Training for data collection. To ensure high-quality and reliable data, data collectors were trained as stipulated by protocols established by observation measure and child assessment developers. For instance, all Classroom Assessment Scoring System (CLASS) observers were

certified as reliable when they demonstrated proficiency of 85 percent or higher (within one point) inter-rater reliability. Training was provided by a certified trainer on the various quality measures. The same data collectors who collected the classroom observations were trained to collect child assessments. Data collectors received rigorous training, and in order to be certified to collect child assessment data were observed by the trainer administering the battery with non-study children. Upon review by a gold standard trainer using a standardized checklist, data collectors were either certified to collect data or went through retraining. Uncertified data collectors were not allowed to collect data. The data collection instruments are described below.

2.2. Measures

The measures used in this study are research-based, used in other QRIS validation studies, and culturally relevant and valid. In Table 2-9, the alignment between standards of ExceleRate Illinois, Circle of Quality Levels, and the measures used in the study are presented. In some instances, there were no agreed upon tools to assess a particular standard (e.g., cultural competence). Thus, we used the best available tools or set of items to independently assess a given standard or domain. Table 2-9 lists independent measures for some criteria (e.g., CLASS scores for standards 1A and 1C on the learning environment and instructional quality; Program Administration Scale (PAS) Curriculum Scale for Standard 1B on curriculum; PAS Staff Qualification items for Standard 13-15 about the director's and classroom staff's qualifications).

Table 2-9. Alignment between ExceleRate Illinois Standards and Criteria and Study Measures

Standard	Criteria	Study Measure
1A Learning Environment	ITERS-R Total/CLASS ES & CO	CLASS ES & CO & ITERS-R for infant classrooms
1B Curriculum	Approved aligned or specific curriculum	Director Interview-list of aligned curricula & PAS #11
1C Instructional Quality	ITERS-R Interactions, Language and Reasoning /CLASS IS	CLASS IS & ITERS-R Listening & Talking, Interactions subscales for infant classrooms
1D Child Screening	Approved screening tools	PAS #10
1E Child Assessment	Approved tools	PAS #11
1F Inclusion	Approved practices	Director Interview
2A Family & Community Engagement	Approved practices	PAS #16 & #17

Standard	Criteria	Study Measure
2B Transition	Approved training	Director Interview-items about plans and implementation
3A Program Administration	Approved tools	Director Interview
3B Group Size & Ratio	Verified by onsite assessor	Classroom observation sheet
3C Continuous Quality Improvement	Verified by onsite assessor	Director Interview & PAS #14
3D Culturally & Linguistically Appropriate Practices	Approved tools	Director Interview
4A Director Qualification	Verified by registry	Director Interview
4B Staff Qualifications	Verified by registry	Director Interview & Teacher Survey
4C Staff Development	Verified by registry	Director Interview

Note. ITERS-R = Infant/Toddler Environment Rating Scale-Revised; CLASS ES = CLASS Emotional Support, CLASS CO = CLASS Classroom Organization, CLASS IS = CLASS Instructional Support; PAS = Program Administration Scale

School and Center Administration. The Program Administration Scale (PAS; Talan & Bloom, 2004) was used to measure program quality across a variety of constructs corresponding to ExceleRate Illinois, including child assessment, child screening, family & community engagement, program administration, quality improvement, and staff development and qualifications. The PAS contains 25 items grouped into 10 subscales that measure the leadership, management, and administrative practices of center-based early childhood programs. For this study, we collected five items: (10) Screening and Identification of Special Needs, (11) Assessment in Support of Learning, (14) Program Evaluation, (16) Family Communications, and (17) Family Support and Involvement. The instrument is designed for multiple uses, including program self-improvement, technical assistance and monitoring, pre-service and in-service training, research and evaluation, and public awareness (Talan & Bloom, 2011). The authors indicated that correlations between the 10 subscales range from .63-.90, inter-rater reliability within 1 point was 90%, and the PAS was moderately correlated (.53) with the ECERS-R, indicating some similarity with an independent quality measure, but capturing unique aspects of program management. A study based on 30 centers in North Carolina found that program administration and organizational climate were both positively correlated with classroom quality, and that level of education of the director was related to higher quality administrative practices (Lower & Cassidy, 2007). Research has yet to link the PAS measure directly to child outcomes. PAS items for this study were collected through a review of materials and interviews with program directors.

Classroom Observations. The CLASS Pre-K Emotional Support and Classroom Organization domains were used to validate standard 1A, learning environment, and the CLASS Pre-K Instructional Support domain was used to validate standard 1C, instructional quality. The Infant/Toddler Environment Rating Scale-R was used to validate learning environment and instructional quality for programs serving infants. In addition, the CLASS and ITERS-R scores were used to determine whether programs in higher Circles of Quality had higher quality in general. During observations of classrooms, the numbers of children and staff were also recorded.

Classroom Assessment Scoring System (CLASS Pre-K; Pianta, LaParo, & Hamre, 2008) is an observational assessment of the quality of teacher-child interactions. Its ten dimensions are organized into three domains. The Emotional Support domain includes positive climate, negative climate, teacher sensitivity, and regard for student perspectives. The Classroom Organization domain includes behavior management, productivity, and instructional learning formats. The Instructional Support domain includes concept development, quality of feedback, and language modeling. Each dimension is rated from 1 to 7 with higher scores indicating higher quality. Data collectors observe classrooms for 4 to 6 cycles of observation for 20 minutes followed by 10 minutes of scoring. Studies have found a link between CLASS domains and other measures of quality, such as the Environment Rating Scales (Early et al., 2006), and with child cognitive and social-emotional outcomes (Burchinal et al., 2008; Mashburn et al., 2008).

Infant/Toddler Environment Rating Scale-Revised (ITERS-R; Harms, Clifford, & Cryer, 2007) is an observational measure used to assess the global quality of child care classrooms serving infants and toddlers up to 30 months. The measure consists of 39 items organized into 7 subscales: (1) Space and Furnishing, (2) Personal Care Routines, (3) Listening and Talking, (4) Activities, (5) Interaction, (6) Program Structure, and (7) Parents and Staff (as is common practice in studies, we did not gather the Parents and Staff subscale). Scores on the ITERS-R range from 1 to 7 with higher scores indicating higher quality. Studies have shown a relationship between the Environment Rating Scales and other indicators of program quality (Early et al., 2006) and with cognitive and social-emotional outcomes (Burchinal et al., 2009; Burchinal et al., 2000). We acknowledge that recent studies have questioned the link between Environment Rating Scales and child outcomes, especially within QRIS (Gordon, Fujimoto, Kaestner, Korenman, & Abner, 2013; Sabol & Pianta, 2014); however, there are limited research-based tools to assess infant classrooms.

Principal/Director interview. During the collection of PAS data, principals and/or program directors were interviewed (depending on who the school or center felt was best to answer questions related to the administration of early childhood classrooms). As needed, a bilingual staff member was on hand to conduct the interview in English or Spanish. The director's interview included questions about the curricula used in the center, any transition plans and their implementation, the ratios and group sizes for each classroom and age of children in that room, quality improvement plans, the extent to which parents who speak a language other

than English had someone who communicated with them in that language, director and classroom staff qualifications in terms of percentages at Illinois credential levels, and staff development plans. In addition, information about teacher turnover and compensation were gathered, factors that have been found to be associated with program quality (Grissmer & Kirby, 1997; Phillipsen et al., 1997).

Teacher survey. A web-based survey was administered to teachers in the selected classrooms. The lead teacher survey included questions about both the teacher/classroom and the selected children in the classroom. Teachers were asked to report on their demographics, including race/ethnicity, education, experience, and tenure, and to provide information for validation of standards including ratio, group size, curriculum, their qualifications, and development plans. In the child outcome portion of the survey, teachers were asked to rate the child's academic and social skills using the instruments described below. Paper copies were made available as needed.

Family Questionnaire. Basic demographic information was collected from parents at the same time that they gave permission for their child to participate in the study. We gathered demographic information about the parent (e.g., age, race/ethnicity, education, income range) and the child (age, gender, race/ethnicity), as well as information about the child's health (see below for more details). The questionnaire was made available in both English and Spanish (based on the Census more than 20% of households in Illinois speak a language other than English, predominantly Spanish).

Child Pre-Academic and Social Skills. We used a multi-method, multi-informant approach for assessing children's learning and development. The measures described below are available in English and Spanish and collectively assess multiple domains of children's development. Two methods were used to determine children's English-language proficiency for child assessments. Children whose parents or teachers indicated that the child's primary language was Spanish were routed to a bilingual version of the assessment. Based on children's performance on an initial set of 15 items from the PreLAS® (Duncan & DeAvilla, 1998), the assessment proceeded in English or the child was routed out of the English version of the assessment and the assessment was given in Spanish. Children whose parents or teachers indicated that the child's home language was a language other than Spanish or English were also given the Pre-LAS. If they passed the initial screening in English then they proceeded with the English assessment. If they did not pass the screening in English, then the assessment did not proceed and another child was assessed instead.

Language and Literacy Skills. We administered three assessments of children's language and literacy skills. First, teachers were asked to complete the Academic Rating Scale (ARS; $\alpha = .95$) from the Early Childhood Longitudinal Study. The ARS gathers information about teachers' perceptions of children's literacy skills. Teachers rate each child's skills on 1= "Not yet"—child has not yet demonstrated skill to 5="Proficient"—child demonstrates skill, knowledge, or behavior competently and consistently." Second, we administered two subtests of the Woodcock-Johnson III Tests of Achievement: Letter-Word Identification and Picture Vocabulary

(WJ-III; Woodcock, McGrew, & Mather, 2001) for English speaking children and the Batería III Woodcock-Muñoz (Muñoz-Sandoval, Woodcock, McGrew, & Mather, 2005) for Spanish-speaking children. Letter-Word Identification measures the basic pre-reading skills of letter and word recognition. Picture Vocabulary measures expressive (speaking) and receptive (understanding) language skills. Large-scale studies using rigorous methods (i.e., IRT) have equated the English and Spanish WJ-III measures and indicate that they assess the same competencies (Woodcock & Munoz-Sandoval, 1993). Recent findings indicate no significant differences between the English or Spanish versions of the WJ-III (Hindman, Skibbe, Miller, & Zimmerman, 2010).

Math Skills. Two measures of children's math skills were used. Teacher ratings on the ARS (α=.94) described the child's early math skills. The WJ III Applied Problems and Batería III Applied Problems subtest measures simple counting, adding, subtracting, and making comparisons for children who speak English and Spanish, respectively.

Executive Function Skills. Two measures were used to assess four-year-old children's executive function skills. The Head-Toes-Knees-Shoulders (HTKS) task is a measure of behavioral self-regulation that requires cognitive flexibility, working memory, and inhibitory control executive function skills (Ponitz, McClelland, Matthews, & Morrison, 2009). The task requires children to engage in a Simon-Says like game that requires children to engage in the opposite behavior of the one described by the assessor. For example, children are asked to "touch your head" and they are given two points for doing the opposite and touching their toes, one point for reaching toward their head and self-correcting, and zero points for an incorrect response. If children pass the head/toes part of the task, they complete an advanced trial where the knees and shoulders commands are added. The task was scored as a total from the 20 test items resulting in a range of possible scores of 0-40. The Pencil Tap task is a measure of the inhibitory control aspect of executive function. The task was adapted for use in the preschool setting (Blair, 2002; Diamond & Taylor, 1996; Smith-Donald, Raver, Hayes, & Richardson, 2007) from a lab-based peg-tapping task first developed by Luria (1966). In the Pencil Tap task, both the assessor and the child have a pencil. Children are instructed to copy the assessor and tap the pencil once or twice on the surface of the table or desk. The assessor then changed the rules and asked the child to tap in a pattern opposite of the assessor's action (e.g., if the assessor taps once, the child should tap twice). Once the child practiced up to 6 times with feedback, the 16 scored trials were administered without feedback. The total number of correct responses was recorded with a range of 0-16. Items exhibited good internal consistency at pre-and post-test assessments (KR-20s = .89 and .91, respectively).

Social-Emotional Skills. The Teacher-Child Rating Scale (T-CRS; Perkins & Hightower, 2002) is a validated teacher-rated measure of students' problem behaviors and competencies, consisting of 32 items assessing four empirically-derived subscales: task orientation, behavior control, assertiveness, and peer social skills. T-CRS alpha coefficients of internal consistency range from .90 to .94.

Approaches to Learning. Approaches to learning were measured using the Preschool Learning Behaviors Scale (PLBS; McDermott, Green, Francis, & Stott, 2000), a teacher-report measure of preschool children's learning behaviors in the classroom. Teachers are asked to rate how often a child exhibits particular behaviors for 29 Likert-type items (2=Most Often Applies, 1=Sometimes Applies, or 0=Doesn't Apply). Four subscales were used in this study: Competence Motivation, Attitude Toward Learning, Attention/Persistence, and Strategy/Flexibility. Convergent and divergent validity for the nationally normed scale have been established (McDermott, Leigh, & Perry, 2002), and this scale has been validated for use with a low-income preschool population (Fantuzzo, Perry, & McDermott, 2004).

Existing Data. We used data about ratings on each domain for every program participating in ExceleRate Illinois to develop the study design for the validation study, and analyzed those existing data to address research questions 1-3.

2.3. Data Analysis

To answer the four research questions posed for this validation study, the analysis plan had four main parts which will be described in detail below:

1. To what extent does ExceleRate Illinois differentiate levels of quality in programs overall and across each domain?

2. To what extent does a program's score on each domain contribute independently to its overall rating?

3. What combination and/or weighting of indicators best discriminate levels of quality?

4. Do children who participate in programs rated at higher Circles of Quality overall and across particular domains demonstrate greater gains in child outcomes compared to children who participate in programs rated at lower levels overall and across domains?

As described above, the study included two sources of data. First, existing data from the past 3 years were shared with the research team by ISBE (for PFA programs) and INCCRRA (for other center-based programs). The ISBE data included overall Circle of Quality ratings as well as information on each of the separate standards within each of the domains. The INCCRRA data just included the overall Circle of Quality ratings. Second, the FPG/AIR team collected validation data on randomly selected programs as described above. Before conducting the analyses addressing the research questions, we examined the ISBE ExceleRate and FPG/AIR-gathered data descriptively. The data analysis plans for addressing each research question are described below.

Research Question 1. The first research question asks whether programs rated at higher Circles of Quality show higher levels of quality overall and within domain. This question was examined using both sources of data. First, we computed domain scores from the standards collected during ExceleRate ratings of programs. As described previously, ExceleRate has four domains:

(1) Teaching & Learning, (2) Family & Community Engagement, (3) Leadership & Management, and (4) Qualifications & Continuing Education. We tested whether programs rated at the Gold Circle scored higher on each domain and on the observed quality measures that were collected during the rating process. T-tests were conducted, comparing the programs rated at the Gold and Silver Circles. We were able to estimate domain scores using data from the ISBE review process, but INCCRRA program reviews do not retain detail about standards or domain scores. As such, only the ISBE data were used to examine domain scores. However, we were able to include observed classroom quality scores from both the INCCRRA and ISBE for analyses related to measures of classroom quality.

Second, we examined data collected by the FPG/AIR team for each of the standards to test whether programs rated at higher Circles of Quality had higher scores on independently assessed measures of quality (i.e., ITERS-R, CLASS Pre-K, and PAS scores). Hierarchical linear model analyses were conducted to test the extent to which centers rated at higher Circles of Quality showed higher classroom quality on the ITERS-R and CLASS Pre-K and center quality on the PAS. Analyses of ITERS-R and CLASS Pre-K scores accounted for the nesting of classrooms in centers, and included region, type of program (PFA vs. other centers and schools), and pathway (i.e., whether the program had been assessed using ExceleRate standards or accredited based on meeting other program standards). Of interest was whether associations between Circles of Quality and independent ratings of classroom and center quality were different based on type of program or pathway (assessed or accredited).

Research Question 2. The second research question asks whether the four domain scores (Teaching & Learning, Family & Community Engagement, Leadership & Management, and Qualifications & Continuing Education) contribute independently to overall ratings. The ISBE ExceleRate ratings of prekindergarten programs were analyzed. A logistic regression analysis predicted Gold Circle of Quality status from the domain scores.

Research Question 3. The third research question asks about the dimensionality of the ExceleRate standards. A factor analysis involving principle components with varimax rotation was conducted using the standards. We were not able to assess whether reweighting the standards provided better prediction of independent quality measures and child outcomes due to the limited number of programs with standards data from ExceleRate and with data collected by the research team.

Research Question 4. The final research question asks whether children's scores on standardized school readiness measures are higher when children attend programs rated at higher Circles of Quality. HLM analyses accounted for the nesting of children in classrooms and classrooms in centers. The analyses included the child's fall score, child and family characteristics, and region as covariates. These "value-added" multi-level analyses included Circle of Quality, pathway, and type of program as the primary predictors and asked whether pathway, type of program, home language, diverse ethnicity, or IEP status moderated associations between Circles of Quality and residualized gains in child outcomes, that is,

whether any of the predictors were more or less related to child outcomes within particular categories of program or for specific groups of children.

Missing Data. Many efforts were made to minimize missing data, but where data were missing we assumed the data were missing at random because few if any of the missing data were a function of the outcome measures (Schafer & Graham, 2002). State-of-the-art, accepted methods were used to handle the missing data.[2]

Section 3. Results

A description of the centers and schools, and child and family outcomes will be followed by a detailed description of the analytic results designed to answer the study research questions.

3.1. Descriptive Statistics for FPG/AIR-Gathered Data

The proposed sampling plan included examining pathway and program type differences among programs at different Circles of Quality, but this was not possible due to the manner in which ExceleRate was rolled out. First, due to the newness of the ExceleRate Illinois system, very few programs had reached the Bronze Circle of Quality so only programs at the Gold, Silver, and Licensed Circles of Quality were sampled. Second, the newness of the system and requirements of quality observations also limited the extent to which programs could be accredited at the Silver Circle of Quality. Third, the absence of Preschool for All programs at the Licensed level limited our ability to examine type of program by Circle of Quality interactions.

Table 3.1 shows the number of classrooms at each Circle of Quality that used the accredited or assessed paths and that were Preschool for All or other center-based programs. As shown below, there was variation in pathway among centers at the Gold Circle. There were too few accredited programs at the Silver Circle to examine the pathway question at the Silver Circle, and it was not possible to examine it at the Licensed Circle because it was not relevant for those programs. Similarly, by design there were no PFA programs at the Licensed Circle.

Table 3.1 Number of Classrooms in Validation Study

	Gold	Silver	Licensed
Pathway			
Accredited	107	6	
Assessed	91	95	
Licensed			58

[2] Fifty imputation data sets were created using missing data imputed from regression analyses of the other variables and adding random error to preserve the degree of variability. Analyses were conducted separately for each data set and then results were combined in a manner that took into account variation within and between imputation data sets (Rubin, 1976, 1987; Schafer & Graham, 2002). We used the Markov Chain Monte Carlo (MCMC) approach that uses multiple chains and completes at least 200 burn-in iterations before each imputation, and uses the Expectation–Maximization (EM) algorithm across iterations. The burn-in iterations are used to make the iterations converge to the stationary distribution before the imputation.

	Gold	Silver	Licensed
Pathway			
School (PFA) vs. Center (Not PFA)			
PFA	63	30	
Not PFA	135	71	58

The 357 classrooms observed in this study varied in quality with infant toddler classrooms generally in the low-moderate quality range, and classrooms serving preschool-age children (3-5 year olds) in the moderate to high range for most domains (see Table 3-2).

Table 3-2. Classroom Quality Scores at the Center/School Level

	Overall (n=153)		
	n	M	SD
ITERS-R Mean Score	79	3.66	0.85
CLASS Pre-K Domains			
Emotional Support	152	5.56	0.70
Classroom Organization	152	5.09	0.83
Instructional Support	152	3.01	1.02
Program Assessment Scale			
10: Special Needs	149	4.36	2.34
11: Assessments	149	5.29	1.98
14: Program Evaluation	149	3.58	2.52
16: Family Communication	149	3.59	2.21
17: Family Support	149	5.09	1.68

Children in preschool-aged classrooms (3-5 years) were assessed in the Fall and Spring, and showed gains over time greater than would be expected based on age-related growth on the following measures: Letter-Word Identification and Applied Problems in English (see Table 3-3). Growth was also observed on teacher-rated social skills and teacher-rated school-readiness, but not teacher-rated learning behaviors. Comparisons on Spanish-language assessments may not be meaningful due to the small number of Spanish-speaking children who participated in the study and differences in who was assessed in the Fall and Spring (only 44 children were assessed in Spanish and of those only 28 had assessments at both time points), so scores on Spanish measures were combined with scores on English language measures in the table below.

Table 3-3. Child Outcome Scores by Time Point and Circle of Quality

	Overall (n=877)					
	Fall			Spring		
	n	M	SD	n	M	SD
Child Assessments – English and Spanish						
Letter-Word Recognition *(standard score)*	713	102.90	13.53	704	104.06	14.04
Picture Vocabulary *(standard score)*	723	102.75	13.84	710	101.75	13.17
Applied Problems *(standard score)*	720	105.01	14.25	703	106.30	14.81
Executive Function HTKS	395	16.03	11.91	490	19.30	12.88
Exec Function Pencil Tap	641	8.39	4.92	655	10.33	4.92
Teacher-Child Rating Scale						
Behavior Control	673	3.57	.82	595	3.63	.89
Assertiveness	673	3.74	.82	595	3.87	.75
Peer Social Skills	673	4.06	.77	595	4.13	.82
Task Orientation	673	3.73	.91	595	3.83	.89
Academic Rating Scale						
General Knowledge	673	3.06	1.02	603	3.58	1.04
Language Arts	670	2.71	.96	602	3.25	1.01
Math	673	2.69	1.04	603	3.29	1.09
Preschool Learning Behaviors Scale						
Attitude Towards Learning	673	1.73	.36	548	1.73	.38
Attention/ Persistence	673	1.51	.50	548	1.53	.51
Competence/ Motivation	673	1.57	.42	548	1.61	.43
Strategy/ Flexibility	673	1.65	.36	548	1.63	.39
Teaching Ratings Combined						
Self-Regulation/ Applied Learning (TCRS & PLBS)	673	4.12	.61	595	4.13	.67
Academic Skills (ARS)	673	2.82	.95	603	3.37	1.00

Overall 15% of parents reported utilizing the Circle of Quality rating in choosing their child's school/center and on average reported being somewhat or very satisfied with: the quality of their child's center/school, their child's school-readiness, relationship with the child's teacher, how much the child learned, and how much the child gets along with their friends at school (1=Not at all satisfied, 2=Somewhat dissatisfied, 3=Somewhat Satisfied, 4 = Very satisfied). Parents reported communicating with their child's teacher/preschool staff weekly and meeting with their child's teacher and receiving the results of screenings and child assessments once or twice a year (scale 1=Never, 2=Once or twice a year, 3=Almost every month, 4=Almost every week, 5=More than once per week). Referral and transition services either were reported as not received or occurring once or twice a year. Almost all parents reported that their center/school made them and their child feel welcome and showed respect for their home

background and child rearing style and values (scale 1 =Disagree, 2=Somewhat Disagree, 3=Somewhat Agree, 4=Agree).

Table 3-4. Family Experiences with the Center/School

	n	M	SD
How satisfied are you with the following?			
Quality of your child's program	591	3.55	.62
How prepared your child is to be successful in school after attending this program	418	3.80	.47
Your child's relationship with his or her teacher	417	3.87	.40
How much your child learned	417	3.82	.45
How much your child gets along with friends at school	417	3.82	.46
Frequency of the following at your child's preschool:			
Communicating with your child's teacher/preschool staff	415	4.24	.94
Had a conference with your child's teacher	414	2.31	.80
Participated in classroom activity	416	2.17	1.07
Received screening test results	411	2.00	.81
Received results from child assessment	413	2.29	.82
Received written plan for referral services	409	1.51	.85
Received support to help child transition into preschool	411	1.95	.99
Received support to help child transition between classrooms	411	1.79	.94
Received support to help child transition to kindergarten	405	1.83	1.00
Do the teachers at your center:			
Make you feel welcome?	419	3.94	.29
Show respect for your home background?	419	3.95	.29
Show respect for your child rearing style and values?	419	3.93	.31
Make your child feel welcome?	419	3.95	.25
Did you consider the ExceleRate Circle of Quality in selecting your preschool? (only answered by parents in rated programs)	194	15%	

Over half of the programs in the FPG/AIR sample had already achieved a Gold Circle of Quality (52%), and 11% reported already applying for a higher circle of quality. Of those who were not already at the Gold Circle and had not already applied, 21% were in the process of applying. Several Directors indicated that finances or administrative resources were a barrier to applying for a higher Circle of Quality. Almost all ECE Directors and Principals reported having Continuous Quality Improvement (CQI) Plans and a majority of staff have Individualized Development Plans (see Table 3-5). A little more than half reported receiving technical assistance, consultation or coaching focused on improving program quality between 3-6 times per year. Almost all reported feeling that the TA/consultation/coaching was helpful or extremely helpful. Despite most programs reporting that they publish and/or advertise their

Circle of Quality rating, they report a lower proportion of families talking with them about their Circle of Quality.

Table 3-5. Director Report of Quality Improvement Activities

	n	%/M	SD
Continuous Quality Improvement (CQI)			
Program has written CQI Plan	128	90%	
Made progress meeting CQI Plan objectives	101	100%	
Individualized Development Plan			
All classroom staff have a professional development (PD) plan	128	82%	
PD plan includes enhancing communication w/families	101	87%	
PD plan includes understanding family cultural practices	101	82%	
PD plan includes support development of children w/special needs	101	87%	
PD plan includes improving children's learning	101	91%	
PD plan includes improving classroom quality	101	96%	
Coaching			
Coach provided assistance to improve center quality	128	57%	
Frequency of coaching in past year? (1=1-2, 2=3-4, 3=5-6, 4=7-8, 5=9-10, 6=10+)	59	2.90	1.68
Helpfulness of assistance? (1=Not at All to 4=Extremely Helpful)	59	3.51	0.63
ExceleRate			
Most parents know program participating in ExceleRate	128	63%	
Publish ExceleRate Circle of Quality	128	84%	
Advertise ExceleRate Circle of Quality	128	73%	
Talked to parents about ExceleRate Circle of Quality	128	59%	
Parent percentage ask about ExceleRate Circle of Quality	128	40%	
Quality Improvement			
Applied for higher ExceleRate Circle of Quality			
Yes	151	11%	
No	151	38%	
Already Gold	151	52%	
If not Gold and didn't apply for higher Circle, reason for not applying:			
Preparing to apply	52	21%	
Staffing qualifications too high	52	15%	
Not enough time	52	8%	
Too much paperwork	52	8%	
Do not know enough about it	52	2%	
Process unclear	52	4%	
Other	52	31%	
How do parents find out about your ExceleRate Rating?			
ExceleRate website	151	2%	

	n	%/M	SD
Another center/school/colleague	151	11%	
INCCRRA	151	16%	
Local CCR&R agency	151	25%	
Do not know about ExceleRate	151	0%	
Other	151	26%	

Almost all ECE Directors and Principals reported that their staff are trained on inclusion of children with special needs and that they provide screenings to children on average once or twice a year (see Table 3-6). Closer to two-thirds have a Memorandum of Understanding with at least one Local Education Agency. Almost three-quarters of programs serve children who are dual language learners and of those, almost all reported that they provide instruction in a language other than English for at least some portion of the day.

Table 3-6. Director Report of Inclusion of Children with Special Needs and Dual Language Learners

	n	%/Mean	SD
Program serves children with special needs	151	95%	
Program administers screenings for disabilities/special needs	151	91%	
How often per year?	137	1.85	0.80
Written policies for IEPs/IFSPs	137	76%	
Written MOU	137	68%	
Written MOU with local CFC	137	59%	
Dual Language Learners			
Program serves dual language learners	151	72%	
Number of classrooms in each center/school offering instruction other than English	108	4.31	3.55
Proportion of classrooms that provide non-English instruction	151	34%	
Spanish only	51	0.22	0.78
Both English and Spanish	51	3.47	2.66
Other language only	51	0.18	0.82
English and other language	51	0.61	1.58
Program translates material into home language	112	75%	
Program has translation service for parents	113	77%	
Staff trained on inclusion of children with special needs	119	95%	
Center/School Administrator	119	34%	
Some Lead Teachers	119	13%	
Some Assistant Teachers	119	17%	
All Lead Teachers	119	28%	
All Assistant Teachers	119	7%	
All classroom teaching staff	119	32%	

Most ECE Directors and/or Principals reported conducting assessments aligned with their curriculum of children's learning and development (see Table 3-7). The majority of programs use Teaching Strategies GOLD/Creative Curriculum, or assessments not on the list of research-based measures. The results of the assessments are reported to be used by almost all centers/schools to inform instruction, and 87% completed training on the assessment tool. Most reported offering all families a variety of parent and family engagement opportunities.

Table 3-7. Director Report of Assessments and Communication with Families

	n	%/Mean	SD
Program administers assessments to assess learning and development	151	93%	
Teaching Strategies GOLD/Creative Curriculum	151	65%	
Work Sampling	151	13%	
High/Scope COR	151	1%	
Early Learning Scales	151	10%	
Other assessments	151	33%	
How often assessed	141	2.61	0.61
How often per year do you meet with parents to discuss assessments?	141	2.30	0.55
Results used to inform instruction	141	99%	
Completed training on assessments	141	87%	
Written Plan in place for all children/families:			
offering daily communication with teacher	151	81%	
offering conference with teacher	151	80%	
offering participation in routine class activities	151	91%	
offering results from a screening test	151	85%	
offering results from an assessment	151	87%	
offering plan to refer families to needed services	151	75%	
offering plan/activities to help transition into preschool	151	82%	
offering plan/activities to help transition between classrooms	151	73%	
offering plan/activities to help transition into kindergarten	151	84%	

3.2. Descriptive Statistics for Existing ExceleRate Data

A Data Use Agreement between Illinois state agencies and UNC was established so ISBE and INCCRRA could share their ExceleRate ratings with us to use in addressing the research questions. We received the 2013, 2014, and 2015 ExceleRate ratings of Preschool for All programs from ISBE and the most recent ratings from community-based organizations, schools, and Head Start programs from INCCRRA. We combined the data, and discovered, however, that only ISBE recorded whether individual standards or indicators were met. Accordingly, only ISBE data could be used when examining the individual standards. We chose to use each center's most recent data. For 33% of the programs, this was 2015, 38% 2014, and 29% 2013.

ExceleRate Quality Assessments. ExceleRate quality assessors collected the Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford, & Cryer, 1998). They use ECERS-R total scores to determine Standard 1A and the ECERS-R Language and Interactions scale scores to determine Standard 1C. To meet Standard 1A at the Gold Circle of Quality, the program must have an across-classroom mean ECERS-R score of 4.5 or above with no classroom scoring below 4. To meet Standard 1C at the Gold Circle of Quality, the program must have a classroom mean ECERS-R Language and Interaction score of 4.75 or above with no classroom scoring below 4.

ExceleRate Domain Scores. Circle of Quality domain scores were computed for the four domains: (1) Teaching & Learning, (2) Family and Community Engagement, (3) Leadership & Management, and (4) Qualifications & Continuing Education. The proportion of Gold standards met was computed overall and within each domain. Table 3-8 shows the individual standards that comprised each domain, and provides descriptive statistics for the standard and domain scores. The internal consistency (Kuder–Richardson Formula 20) was estimated for the scale overall and for each domain. The domain scores for Family and Community Engagement and Qualifications showed good internal consistency, but the Teaching and Learning and Leadership and Management domains did not. If the program met all of the standards required at the Gold Circle of Quality (a score of 1) within a domain, they received a score of 1 for meeting the requirements of that domain at the Gold Circle of Quality.

Table 3-8. Descriptive Statistics for ExceleRate Domain Scores

	n	Mean (sd)	Range	KR20
Overall compliance/proportion of standards met	327	.54 (.27)	0-1	.82
1.Teaching & Learning Domain	327	.60 (.26)	0-1	.49
1A. ECERS-R ≥ 4.5, None <4	443	.67	0-1	
1B. Approved Curriculum	327	.53	0-1	
1C. ECERS-R Language & Interaction ≥4.75, None<4	443	.80	0-1	
1D. Screening (PAS Item)	327	.51	0-1	
1E. Assessment Tools Aligned With Curriculum	327	.59	0-1	
1F. Training On Special Needs	327	.52	0-1	
2.Family & Community Engagement Domain	327	.42 (.46)	0-1	.76
2A. Training On Inclusion	254	.35	0-1	
2B. Written Policies On Engagement (PAS)	327	.24	0-1	
3. Leadership & Management Domain	327	.52 (.31)	0-1	.51
3A. PAS Child Assessment, Family Partnership, Public Relations, and Marketing Scales ≥ 5	327	.32	0-1	
3B. Ratios	327	.60	0-1	
3C. Continuous Improvement	327	.62	0-1	
3D. ECERS-R Diversity Item ≥ 4.5 & Use Home Language	327	.55	0-1	
4.Qualifications & Continuing Education	327	.52 (.50)	0-1	1.0
4A director meets cred level II	327	.52	0-1	

	n	Mean (sd)	Range	KR20
4B 20%+ Teachers met cred level 5	327	.52	0-1	
4C staff dev plans & 20+ hrs PD	327	.52	0-1	
ECERS-R total	443	4.74 (.61)	3.00-6.17	
ECERS-R Language & Interactions	443	5.21 (.513)	3.14-6.59	

For the domains that did not have good internal consistency, we identified the standards that did or did not appear to be moderately to highly correlated with each other. Results are shown below.

- Domain 1: Teaching and Learning (KR20=.49):
  - Standards based on classroom observations using the ECERS-R were highly correlated (indicators 1A and 1C, r=.50)
  - Standards regarding assessments (1D & 1E, r=.30) and regarding special needs (1D & 1F, r=.27) were moderately correlated
  - Other standards were modestly correlated at best.
- Domain 3: Leadership and Management (KR20-.51):
  - Leadership measured with the PAS and continuous improvement standards were highly correlated (3A & 3C, r=.54)
  - Leadership measured with the PAS and ratios standards (3A & 3B, r=.26) were moderately correlated
  - Other standards were modestly correlated at best.

We also examined the ISBE data to determine the degree to which the total number of standards met, the number of standards met within a domain, and ECERS-R total and Language and Interactions scale scores were correlated (Table 3-9). The domain scores were moderately to highly correlated with each other and the ECERS-R total and subscale scores were also highly inter-correlated, but only the Teaching and Learning domain—which includes two standards based on the ECERS-R–was correlated with the ECERS-R scores.

Table 3-9. Correlations among Standards Met within Domains and ECERS-R Scores (n=327-433 centers/schools)

	Family & Community Engagement Domain	Leadership & Management Domain	Qualifications & Continuing Education Domain	Total Standards met	ECERS-R total score	ECERS-R Lang & Interact Scale Score
Teaching & Learning	.35***	.43***	.49***	.79***	.58***	.52***
Family & CommEngagt		.54***	.34***	.64***	.10	.03
Leadership& Management			.47***	.78***	.06	.02
Qualifications & Continuing Ed				.80***	.12	.05

	Family & Community Engagement Domain	Leadership & Management Domain	Qualifications & Continuing Education Domain	Total Standards met	ECERS-R total score	ECERS-R Lang & Interact Scale Score
Total # of Standards met					.36***	.25***
ECERS-R total score						.78***

3.3. Research Question 1: To What extent does ExceleRate Illinois differentiate levels of quality in programs overall and across each domain?

Both the existing data accessed through ISBE and INCCRRA and the data gathered independently by the FPG/AIR research team were used to answer Research Question 1, which had three sub-questions as indicated in the sections that follow.

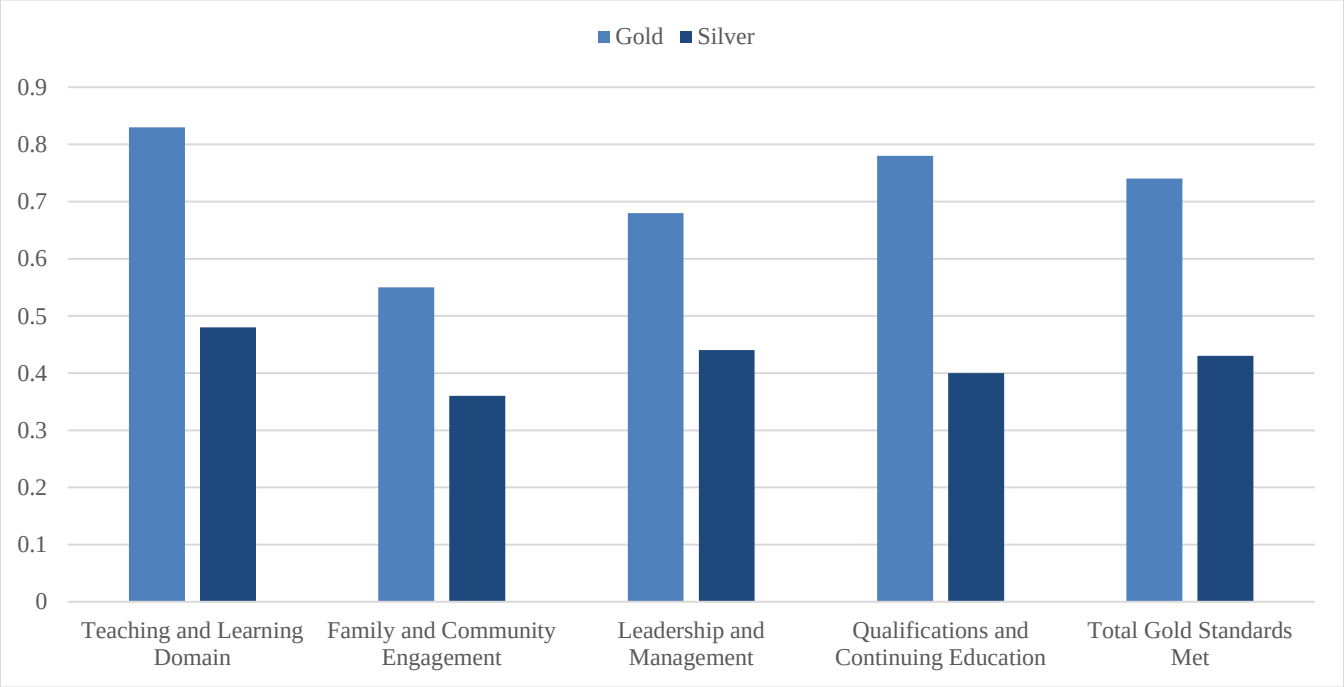
Research Question 1.1. To what extent do programs at higher tiers in ExceleRate Illinois demonstrate higher levels of quality overall and within each domain? Whether and to what extent programs at higher Circles of Quality show higher levels of quality overall or within domain was addressed through analyzing the ISBE ExceleRate data. The programs at the Gold and Silver Circles were compared with t-tests. Results, shown below in Table 3-10 and Figure 3-1, indicate that programs at the Gold Circle had higher quality overall, within domain, and on the ECERS-R total and scale scores. Effect sizes, computed as the difference in the mean Gold and Silver quality scores divided by the pooled SD, were large, ranging from .43 for the Family and Community Engagement domain to 1.85 (i.e., almost 2 standard deviations apart) for the Teaching and Learning domain.

Table 3-10. Comparison of Quality Scores by Circles of Quality

Quality	Gold N M(sd)	Silver N M(sd)	Difference - Effect size	P
Teaching & Learning	113 .83 (.15)	214 .48 (.21)	1.85	***
Family & Community Engage	113 .55 (.46)	214 .36 (.44)	.43	***
Leadership & Management	113 .68 (.25)	214 .44 (.31)	.83	***
Qualifications	113 .76 (.43)	214 .40 (.49)	.77	***
Total compliance	113 .74 (.19)	214 .43 (.24)	1.40	***
ECERS-R total	104 5.13 (.44)	339 4.63 (.56)	.94	***
ECERS-R Lang & Interactions	104 5.50 (.39)	339 5.12 (.51)	.79	***

Note: ISBE data only
*** p<.001

Figure 3-1. Average ExceleRate Domain Scores and Total Standard Scores by Circle of Quality



Note: ISBE data only

Research Question 1.2. According to independent measures of process quality, to what extent do the programs at the highest Circle of Quality demonstrate higher quality compared to other participating programs? Do the programs with Gold ratings have higher quality according to independent measures than other participating programs? To test this research question, we compared the quality scores on the ITERS-R, CLASS Pre-K, and PAS gathered by the FPG/AIR team for mean differences by Circles of Quality.[3] Analyses tested the extent to which the mean scores differed across three Circles (Gold, Silver, and Licensing). All analyses included region and type of program as covariates. Pairwise comparisons of the three Circles of Quality were conducted when the overall comparison was statistically significant. Table 3-11 presents the descriptive statistics (means and standard deviations) for the three Circles of Quality.

Results indicated that there were significant differences between Circles of Quality in CLASS Instructional Support ($F(2,73) = 3.69$, $p < .05$), ITERS-R ($F(2,126) = 7.99$, $p < .001$), and PAS Child Assessment ($F(2,142,) = 2.68$, $p < .05$), Family Engagement ($F(2,142) = 6.09$, $p < .01$), and Program Evaluation Scales ($F(2,142) = 10.38$, $p < .001$) (see Tables 3-12 and 3-13).

[3] The independent assessments of classroom quality were analyzed using hierarchical linear models (HLMs) for CLASS-Pre-K (preschool classroom quality) scores and analyses of covariance (ANCOVA) for ITERS-R scores (infant/toddler classroom quality) and for PAS scores (center quality). The HLMs accounted for the nesting of preschool classrooms within centers and provided the most precise and powerful approach to asking whether classroom quality differed by Circle of Quality. There were so few centers with multiple ITERS-R scores that accounting for nesting in analyses of infant classroom quality was not necessary.

Table 3-11. Average Program-Level Observed Quality by ExceleRate Circle of Quality

	Gold (n=80)			Silver (n=49)			Licensed (n=24)		
	n	M	SD	n	M	SD	n	M	SD
ITERS-R	40	3.81	.8	24	3.82	.86	15	3.00	.66
CLASS Pre-K Domains									
Emotional Support	80	5.53	.69	49	5.66	.66	23	5.48	.83
Classroom Organization	80	5.03	.88	49	5.13	.76	23	5.22	.76
Instructional Support	80	3.2	1.03	49	2.89	1.03	23	2.58	.80
Program Assessment Scale									
10: Special Needs	79	4.33	2.37	47	4.83	2.17	23	3.52	2.39
11: Assessments	79	5.58	1.86	47	5.23	2.03	23	4.39	2.06
14: Program Evaluation	79	4.32	2.48	47	2.49	2.14	23	3.26	2.56
16: Family Communication	79	4.01	2.28	47	2.77	1.95	23	3.83	2.12
17: Family Support	79	5.37	1.61	47	4.98	1.33	23	4.39	2.31

Table 3-12. Differences in Mean Classroom Quality Scores across Three Circles of Quality, Region, and Program Type

		CLASS Emotional Support	CLASS Instructional Support	CLASS Classroom Organization	ITERS-R
# classrooms	N	225	225	225	133
# centers	n	152	152	152	133
Ratings	F	.65	3.18*	.44	9.27***
Gold v Silver	B(se)	-.11 (.12)	.32+ (.19)	-.06 (.14)	-.12 (.18)
Silver v Licensed	B(se)	.18 (.18)	-.21 (.28)	-.12 (.20)	.94*** (.12)
Pre-K	B(se)	-.06 (.12)	.25 (.18)	.17 (.14)	.52 (.64)
Region	F	2.46	.17	7.92***	1.24

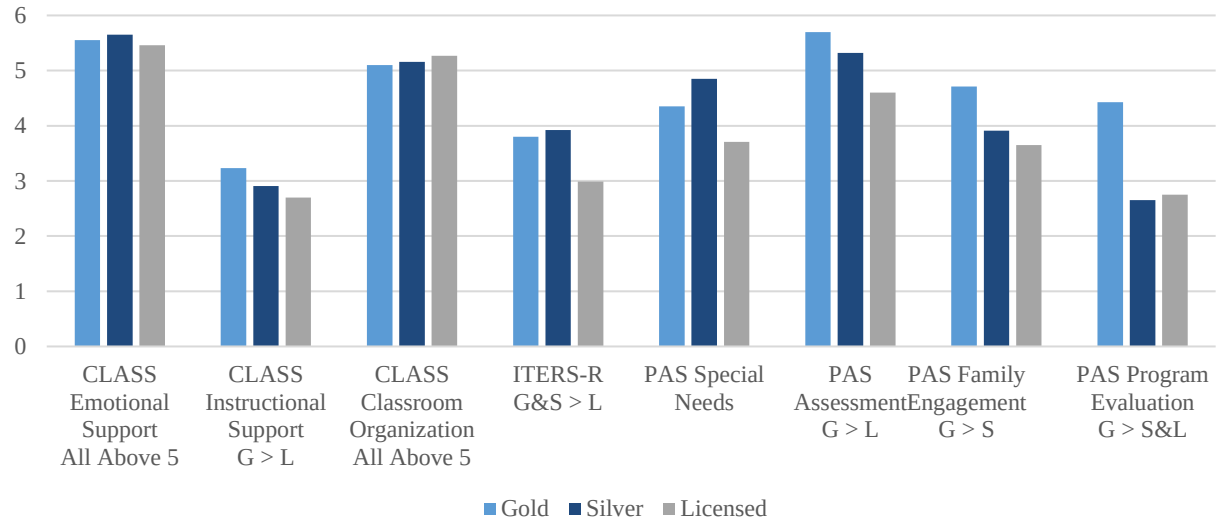
Table 3-13. Differences in Mean Program Quality Scores across Three Circles of Quality, Region, and Program Type

		PAS Identifying Special Needs	PAS Child Assessment	PAS Family Engagement	PAS Program Evaluation
Centers	n	142	142	142	142
Ratings	F	1.79	2.69*	6.09**	10.38***
Gold v Silver	B(se)	-.50 (43)	.38 (.36)	.80** (.28)	1.79*** (.43)
Silver v Licensed	B(se)	1.14^{+} (.61)	.72 (.52)	.26 (.41)	-.10 (.62)
Pre-K	B(se)	.18 (.67)	.35 (.36)	-.96** (.29)	-1.42** (.43)
Region	F	1.62	1.05	3.13*	.50

Average levels of observed classroom and program quality at the center-level by Circle of Quality, controlling for region and program type, are presented in Figure 3-2. Pairwise comparisons of programs at the Gold, Silver, and Licensed Circles of Quality indicated that:

- Centers/Schools rated at the Gold Circle of Quality had higher scores than programs at the Licensing Circle on the following:
  - CLASS Instructional Support (effect size=.48)
  - ITERS-R (effect size=.93)
  - PAS Child Assessment (effect size=.57)
  - Family Engagement (effect size=.69)
  - Program Evaluation Scales (effect size=.73)
- Centers/Schools rated at the Gold Circle of Quality had higher scores than programs at the Silver Circle on the following:
  - PAS Family Engagement (effect size=.52)
  - Program Evaluation Scales (effect size=.77)
  - CLASS Instructional Support (marginal finding, effect size=.29)
- Centers/Schools rated at the Silver Circle of Quality had higher scores than programs at the Licensing Circle on the following:
  - ITERS-R (effect size=1.07)

Figure 3-2. Comparison of Adjusted Means across Classroom and Program Quality Measures by Circle of Quality



Note: Significant Differences are noted below the name of the measure. G=Gold, S=Silver, L=Licensed.

Research Question 1.3: Do associations among Circle of Quality rating and quality as measured independently vary by program type (e.g., Preschool for All compared with other programs) or "pathway" to Circle of Quality level? To answer this question, we examined the interaction between type of program (Preschool for All vs. non-PFA centers and schools) and whether the Circle of Quality was determined based on direct assessment or assumptions made related to program accreditation and the Silver and Gold Circles of Quality. As discussed previously, it was not possible to totally cross Circle of Quality with pathway or type of program because of limitations in the system and data, but we were able to conduct some comparisons to address these questions. We created two interaction terms:

- Comparison of accredited and assessed among Gold Circle centers
- Interaction between Gold/Silver and PFA/not PFA

The first term (accredited or assessed) was included in all analyses to examine whether there were associations between quality and pathway, and the second term (PFA program or not) was included in all analyses of preschool classroom quality or overall center quality (i.e., it could not be included in analyses of the ITERS-R because that was collected only in "not PFA" programs).

Results are shown below in Tables 3-14 and 3-15. Two-level HLM analyses of CLASS and ITERS-R were conducted as before to account for nesting of classrooms in centers. Results indicated:

- No evidence of differences in quality among the PFA and other programs at the Gold Circle of Quality.

- No evidence of differences in quality based on pathway to either the Silver or Gold Circle of Quality.
- PFA programs were rated lower on the PAS Family Engagement and Program Evaluation Scales compared to other programs.

Table 3-14. Differences in Mean Classroom Quality Scores across Three Circles of Quality, Pathway, and Program Type

		CLASS Emotional Support	CLASS Instructional Support	CLASS Classroom Organization	ITERS-R
Classrooms	n	225	225	225	133
Centers	n	152	152	152	133
Ratings	F	1.07	3.17*	0.14	10.34***
Gold v Silver	B(se)	-0.12 (12)	0.36^{+} (.20)	0.03 (.14)	-0.02 (.19)
Silver v Licensed	B(se)	0.27 (.19)	0.14 (.27)	-0.11 (.21)	0.95*** (.23)
PFA	B(se)	-0.25 (.17)	-0.40 (.37)	0.11 (19)	0.25 (.66)
PFA x Gold/Silver	B(se)	0.03 (.16)	0.21 (.24)	0.23 (.19)	
Pathway: Gold	B(se)	0.35 (.25)	-0.20 (.30)	-0.02 (.29)	0.40 (.26)
Region	F	1.83	0.79	7.04***	1.66

Table 3-15. Differences in Mean Program Quality Scores across Three Circles of Quality, Pathway, and Program Type

		PAS Identifying Special Needs	PAS Child Assessment	PAS Family Engagement	PAS Program Evaluation
Centers	n	148	148	148	148
Ratings	F	2.23	2.58^{+}	4.81**	9.83***
Gold v Silver	B(se)	-0.65 (44)	0.34 (.37)	0.69* (.29)	1.72*** (.44)
Silver v Licensed	B(se)	1.29* (.64)	0.78 (.54)	0.34 (.42)	0.09 (.65)
PFA	B(se)	0.06 (.57)	0.24 (.49)	-1.02** (.38)	-1.86** (58)
PFA x Gold/Silver	B(se)	-0.76 (.55)	-0.20 (.47)	-0.68^{+} (.37)	-0.23 (.56)
Pathway: Gold	B(se)	0.82 (.87)	0.31 (.74)	0.45 (.57)	0.94 (.88)
Region	F	1.69	1.07	3.10*	0.55

3.4. Research Question 2: To what extent does a program's score on each domain contribute independently to its overall rating?

This question was addressed using the ISBE ExceleRate data and had a two-pronged sub-question as listed below. Because the analyses were conducted with ISBE data only, the findings do not represent all programs in the system and should be interpreted cautiously.

Research Question 2.1. Which quality domains (e.g., Teaching & Learning, Family & Community Engagement) and subdomains (e.g., Learning Environment, Curriculum) best predict ExceleRate Circle of Quality rating? Is the number of standards met within a given domain higher in programs at higher Circles of Quality for each domain? This question was addressed using the ISBE ExceleRate data. A logistic regression examined which quality domain scores provided statistically significant prediction of a Gold Circle of Quality rating using the ISBE PFA data. Effect sizes were computed as odds ratios by taking the exponent of the parameter estimates, which are log odds ratios in logistic regressions. A second logistic regression was conducted which also included the total ECERS-R and ECERS-R Language and Interactions Factor scores as predictors to ensure that findings regarding the Teaching and Learning domain were not due to the ECERS-R scores included in that quality domain. We were able to conduct these analyses using only checklist data from schools that received their quality ratings through ISBE since the data collected on individual standards for ratings conducted by INCCRRA are not stored in a centralized database. Because PFA program requirements include the provision that all teachers must have a Professional Educator License with Early Childhood Endorsement, all teachers have Bachelor degrees and therefore there is little variance in qualifications. This limited these analyses regarding the staff qualifications domain in particular.

Results indicated that a Gold Circle of Quality rating was highly predicted from the domain scores, $X^2(4) = 207.0$, $p<.001$. The Teaching and Learning domain (Odds Ratio > 10) and Leadership and Management domain (Odds Ratio=7.88) were the strongest positive predictors of a Gold Circle rating. In contrast, after accounting for the other domains, Family and Community Engagement was a negative predictor (Odds Ratio=.20), that is, programs with Gold Circle ratings had significantly lower scores on Family and Community Engagement than non-Gold programs. These results largely were unchanged when the ECERS-R scores were added. The overall model, $X^2(6)=217.5$, $p<001$, indicated the domain and ECERS-R scores were good predictors of Gold Circle status, that is, Teaching and Learning, Leadership and Management, and ECERS-R scores were positive predictors and Family Engagement domain was a negative predictor.

These analyses with data on PFA programs suggest that Teaching and Learning and Leadership and Management appear to be the best predictors of the Gold Circle rating, with Family and Community Engagement and Qualifications and Continuing Education providing little additional prediction of the overall Gold Circle rating; however, as noted above, the staff qualifications domain had little variance because of PFA requirements for Bachelor degrees. This means we were limited in our ability to detect associations between the Qualifications and Continuing Education domain and overall ratings.

Table 3-16. Associations between ExceleRate Domain Scores and the Likelihood of Receiving a Gold Circle of Quality Rating

	Intercept	1. Teaching & Learning	2. Family & Community Engagement	3. Leadership & Management	4. Quals & Cont Ed	ECERS-R	ECERS-R Language/ Interactions
Model 1							
B	-9.16	12.01***	-1.63**	2.28**	-0.10		
(se)	(1.06)	(1.51)	(.51)	(1.19)	(.39)		
Odds ratio		>10	0.20	7.88	0.83		
Model 2							
B	-20.42	10.37***	-1.34*	2.60**	0.11	1.18*	1.11+
(se)	(3.22)	(1.64)	(.53)	(.85)	(.42)	(.53)	(.59)
Odds ratio		>10	0.26	13.50	1.12	3.26	3.03

3.5. Research Question 3: What combination and/or weighting of indicators best discriminate levels of quality?

There were four sub-questions for this research question, and we were limited in our ability to answer the questions because of having few programs in the study sample that had both FPG/AIR-gathered data and existing ISBE data that contained information about individual standards within domains.

Research Question 3.1. To what extent does each of the indicators contribute to the overall domain rating? Are there indicators that appear to be more strongly related to the overall domain score computed without that indicator or to independent measures of process quality?
Research Question 3.2. Is there evidence that the indicators might be multidimensional, suggesting that combining them with other indicators within that domain could be problematic in terms of a continuous quality improvement model?
To answer both of these questions, we examined the psychometric properties of the rating scale. With the descriptive analyses in Section 3.2, we reported that the domain scores for Family and Community Engagement and Qualifications showed good internal consistency, but the Teaching and Learning and Leadership and Management domains did not.

To follow-up these descriptive analyses, a factor analysis with varimax rotation was conducted to examine the extent to which the four domains emerged empirically using the checklist data from ISBE. Table 3-17 reports indicators with factor loadings of .30 or greater. Results show good support for the Qualifications domains and some support for Teaching and Learning and Family and Community Engagement domains, but almost no support for the Leadership domain as a unidimensional construct. However, as shown in the table, the standards load on factors that cross the conceptual domains of ExceleRate, suggesting that different combinations of standards might be warranted. Because the analyses were conducted with ISBE data only, the findings should be interpreted cautiously.

Table 3-17. Factor Loadings for ISBE Checklist Items by Quality Domain

	Factor 1	Factor 2	Factor 3	Factor 4
Teaching & Learning Domain				
1A. ECERS-R ≥ 4.5, None <4			.84	
1B. Approved Curriculum			.37	
1C. ECERS-R Language & Interaction ≥4.75, None<4			.76	
1D. Screening (PAS Item)		.89		
1E Assessment Tools Aligned with Curriculum		.49	.39	
1F Training on Special Needs	.98			
Family & Community Engagement Domain				
2A. Training on Inclusion		.77		
2B. Written Policies on Engagement (PAS)		.84		
Leadership & Management Domain				
3A. Selected PAS Scales ≥5	.74			
3B. Ratios		.92		
3C. Continuous Improvement				.54
3D ECERS-R Diversity Item ≥ 4.5, None <4				.82
Qualifications & Continuing Education				
4A. Director Meets Credential Level II	.98			
4B. 20%+ Teachers met Credential Level 5	.98			
4C. Staff Dev Plans & 20+ Hrs PD	.98			

Research Question 3.3. Do we see evidence that the domains might be multidimensional, suggesting the need for more than one total score (e.g., Family & Community Engagement might be important, but may not be related to classroom structural and process quality)?
There were 80 centers for which we had both existing ExceleRate and FPG/AIR-gathered data. We could only calculate the standards for the 22 (13 Gold and 9 Silver) ExceleRate centers for which we had detailed ISBE monitoring data (checklists and ECERS-R data). Because of the small sample size, the analyses we conducted to examine this research question must be considered exploratory and results interpreted with caution.

With the sample of 22 programs with sufficient data, we computed means for the classroom and program quality measures and fall-spring changes in child outcomes and estimated correlations between the ExceleRate domains and the data gathered by the research team. Results are shown below in Tables 3-18 and 3-19. Correlations ranged from small to large, but due to the small sample size no correlations reached levels of statistical significance. Largest correlations were observed between the PAS Special Needs item and ExceleRate domains 2 (Family and Community Engagement) and 4 (Qualifications and Continuing Education). Large negative correlations were observed between the PAS Program Evaluation item and the total number of Gold standards met as well as domains 2 (Family and Community Engagement) and 3 (Leadership and Management). Changes in executive functioning as measured by the Pencil

Tap measure had small positive associations with domains 2 (Family and Community Engagement) and 3 (Leadership and Management). Early literacy skills were negatively associated, in the small to moderate range, with total number of Gold standards met and with domains 2 (Family and Community Engagement) and 4 (Qualifications and Continuing Education); literacy skills also had small positive associations with environmental quality scores. A similar pattern was observed for the self-regulation composite, with a smaller magnitude of associations. Correlations with changes in the other child outcome domains were in the small range. As noted above, these exploratory analyses should be interpreted cautiously, but the findings may be of use in considering revisions to the system.

Table 3-18. Correlations between Classroom and Program Quality Measures and ExceleRate Rating Checklist Standards, Domains, and Quality Assessments

	Total # Gold standards met	Teaching & Learning Domain	Family & Community Engagement Domain	Leadership & Management Domain	Qualifications & Cont. Ed Domain	ECERS-R total	ECERS-R Language & Interactions
CLASS Pre-K							
Emotional Support	.03	.06	-.15	.03	-.08	.09	-.03
Classroom Organization	-.18	-.10	-.24	-.09	-.14	.01	-.19
Instructional Support	-.23	-.01	-.23	-.24	-.25	-.12	-.09
PAS							
Special Needs	.27	.04	.36	.42	.08	-.30	-.27
Child Assessment	-.17	.12	-.37	-.16	-.07	.29	.06
Family Engagement	-.23	-.05	-.20	-.21	-.14	-.06	.11
Program Evaluation	-.57	.30	-.59	-.55	-.36	.14	.25

Note: None of the Pre-K programs had infant/toddler rooms, so they do not have ITERS-R data. CLASS Pre-K = Classroom Assessment Scoring System Pre-K Version, PAS = Program Administration Scale.

Table 3-19. Correlations between changes in Child Outcome Scores and ExceleRate Rating Checklist Standards, Domains, and Quality Observations

	Total Gold standards met	Teaching & Learning Domain	Family & Community Engagement Domain	Leadership & Management Domain	Qualifications & Cont. Ed Domain	ECERS-R total	ECERS-R Language & Interactions
HTKS S-F	.12	.02	-.05	.04	.27	.-.34	-.12
Pencil Tap S-F	.13	.00	.19	.19	.04	-.11	.06
WJ L-Word S-F	-.22	.06	-.21	-.15	-.35	.20	.22
WJ ApProb S-F	.04	-.15	.13	.-.12	.05	-.09	.02
WJ Pic Vocab S-F	-.09	.05	.02	-.12	-.19	-.08	-.15
Self-Regulation	-.04	-.21	-.22	-.01	.22	.11	.09
Academic skills	.09	.11	.03	.04	.08	.17	.04

Note: HTKS = Head Toes Knees Shoulder Assessment, WJ = Woodcock-Johnson, L-Word = Letter-Word Identification (Literacy), ApProb = Applied Problems (Math), Pic Vocab = Picture Vocabulary (Receptive Language), S-F = Fall scores subtracted from Spring scores.

Research Question 3.4. Does weighting the domains and indicators produce total scores that are more strongly related to independent quality measures and child outcomes? We were unable to answer this question because of limited data.

3.6. Research Question 4. Do children who participate in programs rated at higher Circles of Quality overall and across particular domains demonstrate greater gains in child outcomes compared to children who participate in programs rated at lower levels overall and across domains?

The final set of analyses asked whether children showed larger gains in academic and social skills if they attended programs rated at higher Circles of Quality. Three sub-questions were addressed, examining prediction of children's skills by Circles as well as examining special populations, program type, and pathway.

Research Question 4.1. Adjusting for family characteristics, to what extent are the gains in children's academic and social skills larger among children who attended programs rated at higher Circle of Quality levels compared to children in programs rated at lower levels? This research question examined whether children's residualized gain scores were larger when they attended programs with higher Circles of Quality. HLMs analyzed the spring scores on selected outcomes using the child's fall score on that measure as a covariate to account for individual differences when children entered the program. In addition, the child's gender, race/ethnicity, special education status, whether the family spoke a language other than English at home, poverty status, maternal education, region, and type of program were included as covariates.

The analyses accounted for the nesting of children in centers.[4] All of the data were standardized to have a mean of 0 and SD of 1 for all variables, and thus model coefficients can be interpreted as effect sizes.

The child outcomes included measures of academic skills (Woodcock Johnson Letter-Word Identification, Applied Problems, and Picture Vocabulary – with the parallel measure Batería administered to children who spoke Spanish), and executive functioning (Head-Toes-Knees-Shoulders and Pencil Tap). In addition, teachers rated social skills (Teacher-Child Rating Assertiveness, Behavior Control, Peer Social Skills, and Task Orientation Scales), approaches to learning (PBLS Attitudes toward Learning, Attention/Persistence, Competence/Motivation, and Strategy/Flexibility Scales), and academic skills (Academic Rating General Knowledge, Language Arts, and Math Scales). The English and Spanish assessments on the Woodcock Johnson and Batería were combined and a dummy variable indicated whether the child had been tested in English or Spanish. There were too few children tested in Spanish to warrant separate analyses of their data (only 3% of the sample had two assessments in Spanish). A factor analysis of the teacher ratings yielded two factors that accounted for over 70% of the variance in the fall and spring ratings. The first factor consisted of the TCRS and PBLS scale scores and the second factor consisted of the ARS scales. Two summary scores were computed as means of the scale scores after rescaling the PLBS to the 1-5 range of the TCRS and reverse scoring the negative scales. The alphas for the resulting summary scores ranged from .92 to .96. Average child outcome scores by Circle of Quality are presented in Table 3-20.

Table 3-20. Child Outcomes, Fall and Spring, by ExceleRate Circle of Quality

	Gold			*Silver*			*Licensed*		
	N	*Mean*	*Sd*	*N*	*Mean*	*Sd*	*N*	*Mean*	*Sd*
Covariates									
Maternal education	410	3.95	1.71	202	3.50	1.50	104	4.14	1.46
Gender (Male=1)	456	0.47	0.50	221	0.50	0.50	115	0.49	0.50
Non-English language spoken in home	421	0.35	0.48	209	0.25	0.44	109	0.22	0.42
Child has an IEP/IFSP	491	0.09	0.28	244	0.07	0.26	138	0.04	0.19
Minority race	422	0.44	0.50	210	0.44	0.50	109	0.29	0.46
Child has an IEP/IFSP	491	0.09	0.28	244	0.07	0.26	138	0.04	0.19
Age at Spring testing	392	55.69	7.41	199	57.91	6.82	100	54.66	8.73
Child outcomes – fall and spring									
HTKS total fall	372	8.95	11.80	192	9.01	11.68	121	10.50	13.18
HTKS total spring	389	13.26	13.59	197	14.25	14.20	98	15.20	14.84
Pencil Tap total fall	372	7.91	5.16	193	7.85	5.24	117	7.78	5.15
Pencil Tap total spring	391	9.50	5.37	197	10.58	5.07	97	9.90	5.31

[4] 3-level HLMs were proposed but preliminary analyses indicated there was too little variation among classrooms within the programs to estimate random classroom intercepts.

	Gold			Silver			Licensed		
	N	*Mean*	*Sd*	*N*	*Mean*	*Sd*	*N*	*Mean*	*Sd*
WJ Letter Word Id fall	388	103.2	14.24	199	101.0	12.11	126	104.9	13.13
WJ Letter Word Id spring	402	103.9	15.38	202	103.1	11.83	100	106.8	12.13
WJ Applied Problems fall	391	104.8	15.03	203	103.5	12.94	126	108.2	13.36
WJ Applied Problems spring	403	105.6	15.54	200	105.7	12.92	100	110.3	14.89
WJ Picture Vocabulary fall	393	102.5	14.35	204	100.9	13.86	126	106.5	11.34
WJ Picture Vocabulary spring	406	101.3	13.73	204	101.1	12.42	100	105.2	11.86
Self-Regulation Composite fall	363	4.08	0.63	194	4.12	0.60	116	4.23	0.57
Self-Regulation Composite spring	346	4.13	0.68	167	4.12	0.71	78	4.18	0.59
ARS Academic Skills fall	363	2.76	0.94	194	2.92	0.94	116	2.82	0.98
ARS Academic Skills spring	349	3.33	1.00	169	3.49	0.98	81	3.36	1.04

Analyses also accounted for missing data on child and family covariates and child outcomes using appropriate techniques.[5]

Results are presented in the rows labeled Model 1 in Table 3-21. There was no evidence that children in centers rated at Gold or Silver Circles of Quality showed larger gains on the selected study measures during the study period (~5 months). This finding is discussed in Section 4 below.

Table 3-21. HLM Analyses of Child Outcomes by ExceleRate Circles of Quality

Rating Level	HKTS (Executive Functioning)	Pencil Tap (Executive Functioning)	WJ/B Letter Word (Reading)	WJ/B Applied Problem (Math)	WJ/B Picture Vocabulary (Language)	TCRS/PBS Self-Regulation	ARS academics
	B(se)	B(se)	B(se)	B(se)	B(se)	B(se)	B(se)
Model 1							
Quality Rating							
Gold	0.05(0.07)	0.09(0.08)	0.07(0.06)	0.02(0.05)	0.06(0.05)	0.04(0.07)	0.06(0.08)
Silver	0.08(0.10)	-0.012(0.12)	0.02(0.09)	0.02(0.08)	0.07(0.08)	0.14(0.09)	0.01(0.11)

[5] Multiple imputations were conducted to account for missing data using Rubin's approach. Each variable was regressed on all other variables included in any of the analyses (classroom quality, fall and spring child outcomes, and family and child characteristics), and predicted values were computed from that regression for all missing values. This process was completed for all variables, and then the variances and covariances among them was updated in a manner that added some random variability to ensure the variability was not reduced due to imputation. The cycles of imputing missing values and updating the variance matrix repeats until changes across iterations are very small. In all, 50 imputation datasets were created and all analyses were conducted with each data set. Results were combined across data sets by computing across analyses of each imputed dataset the mean of each coefficient and combined variability within and between datasets of each coefficient.

Rating Level	HKTS (Executive Functioning)	Pencil Tap (Executive Functioning)	WJ/B Letter Word (Reading)	WJ/B Applied Problem (Math)	WJ/B Picture Vocabulary (Language)	TCRS/PBS Self-Regulation	ARS academics
	B(se)	B(se)	B(se)	B(se)	B(se)	B(se)	B(se)
Pre-K	-0.06(0.07)	-0.051(0.09)	-0.07(0.06)	-0.01(0.06)	0.00(0.06)	-0.08(0.06)	-0.17(0.08)*
Model 2							
DLL x Gold	-0.09(0.15)	0.005(0.17)	0.02(0.12)	-0.10(0.12)	0.04(0.12)	0.10(0.12)	-0.02(0.13)
DLL x Silver	0.06(0.22)	0.089(0.26)	0.10(0.17)	-0.12(0.17)	0.14(0.17)	-0.07(0.18)	0.03(0.18)
Model 3							
Race x Gold	-0.19(0.13)	-0.007(0.16)	0.00(0.11)	-0.05(0.11)	0.00(0.11)	-0.03(0.12)	0.01(0.13)
Race x Silver	-0.12(0.21)	-0.071(0.23)	0.10(0.15)	-0.03(0.16)	0.03(0.16)	-0.15(0.17)	0.03(0.19)
Model 4							
IEP x Gold	0.21(0.27)	-0.16(0.28)	-0.04(0.19)	0.20(0.18)	-0.14(0.20)	0.08(0.20)	-0.02(0.19)
IEP x Silver	0.34(0.42)	-0.091(0.45)	0.01(0.32)	0.01(0.31)	-0.18(0.34)	0.10(0.35)	0.08(0.32)
Model 5							
Pre-K x Rating	-0.10(0.18)	-0.193(0.22)	-0.03(0.15)	-0.12(0.14)	-0.11(0.14)	-0.03(0.16)	-0.02(0.20)
Pathway: Gold	-0.11(0.15)	0.102(0.18)	0.03(0.12)	0.12(0.11)	0.03(0.11)	-0.15(0.14)	-0.07(0.17)

Research Question 4.1a. To what extent does each domain predict gains in child outcomes? and Research Question 4.1b. To what extent does each standard predict gains in child outcomes?
The analysis plan involved merging the existing ExceleRate data on individual domains and standards with the FPG/AIR-gathered data to address these sub-questions. However, as noted previously, there was a lack of overlap between the two data sources. There were only 22 centers that were assessed by ISBE containing detailed information about standards and included in our sample for data collection. INCCRRA did not code individual standards in a manner that was useful in addressing this issue. We were not able to explore these research questions.

Research Question 4.2. When children attend programs rated at higher Circles of Quality, are gains larger for children from low-income families, ethnic minority children (e.g., African American, Latino), from families in which English is a second language, or who have special needs than for other children? Whether gains were larger when children attended programs rated at higher Circles of Quality for subpopulations of children was tested for three child characteristics: home language, diverse race/ethnicity (African American, Native American, or Hispanic), and IEP status. Separate models tested whether each of these characteristics interacted with Gold or Silver Circle of Quality in predicting outcomes. Results are shown in Table 3-21 in the rows marked as Model 2 for home language interactions, Model 3 for diverse race/ethnicity interactions, and Model 4 for IEP interactions. Because none of these characteristics interacted with Circle of Quality in predicting child outcomes, there was no evidence of larger gains for children from these subpopulations (i.e., there was no evidence of

moderation of the association between Circle of Quality and child outcomes for these subpopulations).

Research Question 4.3. Do child gains related to attending programs rated at higher Circles of Quality vary depending on type of program (e.g., licensed child care setting, Head Start, Preschool for All) or "pathway" to Circle of Quality level? Whether associations between Circles of Quality and gains in children's outcomes differed on the basis of pathway or type of program was tested by adding two more predictors to the Circles of Quality-child outcomes model presented in Table 3-21. In the row marked Model 5, two variables were added: (1) an interaction variable that tested whether differences in gains in child outcomes between children in programs rated at Gold and Silver Circles was larger or smaller if that program was a Preschool for All or center program and (2) a variable indicating whether Gold programs were accredited or assessed. The results shown in Table 3-21 in the rows marked as Model 5 show that there was no evidence of differences in associations between gains and Circles based on either program type or pathway.

Section 4. Summary and Conclusions

ExceleRate Illinois is a statewide quality recognition and improvement system designed to establish standards for program quality identified as important for families and children. The wider literature on discrete aspects of early care and education quality have extensively examined the efficacy of early care and education as an intervention for young children, so the focus of this study was on validating the particular design of ExceleRate to compile those aspects of quality into a single, overarching Circle of Quality rating. The results of this study are designed to provide information to guide future refinements in the design of ExceleRate Illinois. To accomplish this, we conducted a multi-pronged validation of ExceleRate Illinois (Zellman & Fiene, 2012), which included examining associations with independent quality measures; evaluating the properties of the rating, such as associations among domains and standards; and assessing the association between Circles of Quality and gains in preschool-age outcomes. This section presents a summary of the key findings, limitations of the study, and implications and conclusions.

4.1. Key Findings

To what extent does ExceleRate Illinois differentiate levels of quality in programs overall and across each domain?

The ExceleRate Circles of Quality successfully distinguished between classroom and program quality, so that centers and schools at higher Circles of Quality had higher quality scores than schools and centers at lower Circles, on both external validated quality measures and on internal quality domain scores. These results reflect the close alignment between the design of the ExceleRate QRIS and the research literature on child care quality (see Burchinal, Zaslow, & Tarullo, 2016 for a review). Differences by Circle of Quality emerged for six of the eight quality

constructs used for validation by the FPG/AIR team, including the quality of preschool classroom instructional support, infant-toddler classroom global quality, and center practices regarding child assessment, family engagement, and program evaluation. The other two quality constructs, preschool classroom emotional support and classroom organization, were in the high range across Circles of Quality. In addition to external validation, it was reassuring that programs rated as Gold had higher quality domains within the ExceleRate rating system than programs rated at the Silver Circle.

ExceleRate Illinois is also designed to streamline with reporting mechanisms required for centers and schools to complete in fulfillment of their funding, reporting, and/or accreditation requirements. As such, centers and schools may meet ExceleRate standards by undergoing a full ExceleRate review or receive credit for certain ExceleRate standards through alternate pathways. Encouragingly, the results of this study show no differences related to the path through which programs achieve their Circle of Quality rating suggesting that the criteria have been sufficiently aligned across systems at the highest quality tier – the Gold Circle.

To what extent does a program's score on each domain contribute independently to their rating?

Because the analyses were conducted with ISBE data only, the findings related to domain scores should be interpreted cautiously. Using the data provided to the study team, evidence suggests that the Teaching and Learning and Leadership and Management domains were the most strongly predictive of the overall Circle of Quality rating. Neither the Family and Community Engagement nor Qualifications and Continuing Education domains added to the prediction of the Circle of Quality ratings. However, PFA program requirements related to teacher qualifications (all are required to have BAs) limited the variability of domain scores in the ISBE sample and thus limited the ability of the study to find associations between staff qualifications and Circle of Quality ratings.

Additionally, the ExceleRate rating system appears to be multidimensional, and thus it is not surprising that not all domains equally contribute to a single quality rating. This finding is consistent with growing concerns regarding creating a single quality rating from a multidimensional scale. A meta-analysis conducted with data from six large studies of child care quality suggested quality is multidimensional, and that selected classroom quality and leadership characteristics were related both to overall quality and child outcomes, but that not all widely used indicators were (Burchinal et al., 2016).

What combination and/or weighting of indicators best discriminates levels of quality?

We were not able to fully answer questions about weightings because of limited data. We did, however, conduct a factor analysis that suggested the exploration of a different grouping of quality indicators might be warranted. The following structure emerged from this analysis:

- Classroom Quality: Learning Environment (1A), Curriculum (1B), Instructional Quality (1C), and Child Assessment (1E)

- Structural Quality: Child screenings (1D), child assessments (1E), Family and Community Engagement (2A), Transitions (2B), and Group Size and Staff/Child Ratios (3B)
- Program Administration: Program Administration (3A) and Inclusion of Children with Special Needs (1F)
- Quality Improvement: Continuous Quality Improvement (3C) and Culturally and Linguistically Appropriate Practice (3D)

ExceleRate administrators may want to further explore this structure as they consider revisions to the system.

Do children who participate in programs rated at higher Circles of Quality overall and across particular domains demonstrate greater gains in child outcomes compared to children who participate in programs rated at lower levels overall and across domains, and do findings differ by pathway through which the program entered ExceleRate Illinois?

Children's scores on measures were well within what would be expected of a general population, and children in the study evidenced gains from fall to spring (~5 months) that were generally larger than what would be expected based growth related to age. Adjusting for family characteristics, the chosen instruments did not detect gains in children's academic and social skills that were related to Circles of Quality over the short time frame of the study. In addition, associations between child outcomes and Circles of Quality did not differ by home language, diversity in race and ethnicity, or the child's special needs status. Associations between child outcomes and Circles of Quality also did not differ depending program type or pathway.

As indicated above, the overall Circle of Quality ratings reflect a multidimensional construct designed to relate to varying aspects of children's development, learning, well-being, and family functioning. It is therefore not a surprise that when these disparate constructs were combined and related to child outcomes, no statistically significant associations emerged.

Because of limited data, we were unable to conduct analyses to examine the associations between domains and individual standards and child outcomes to examine whether discrete aspects of the ExceleRate ratings were related to child outcomes. The study showed that the overall rating was not associated with outcomes, but we could not examine whether the disparate constructs that make up the rating are related to outcomes. Future evaluations of ExceleRate Illinois should ensure that the data are available to fully examine associations among individual standards, domains, and overall ratings and among domains and children's outcomes.

4.2. Study Limitations

This validation study had several limitations that must be considered. Limitations relate to timing, sampling, and contextual issues.

Timing. Two limitations related to timing are likely related to study findings. First, the study was conducted as the ExceleRate system was being rolled out, and thus evaluation activities

were occurring before the system was fully implemented. We know from the field of implementation science that it takes 2-4 years for new interventions to be fully implemented (Fixsen et al., 2005); the validation study was launched during very initial implementation stages of ExceleRate, with sampling for the study occurring as programs were just entering the system. Evaluating an intervention before it has reached full implementation may not yield valid findings about the functioning of the intervention, in this case, the ExceleRate system. Second, the fall-to-spring study period for child assessments averaged just under 5 months, which is not unusual for research in early care and education settings. The fall data collection point was somewhat later, on average, than planned because of challenges in program recruitment related to contextual factors (see section below). A 5-month window does not allow much time for children to evidence significant growth in the domains of development that were assessed in the study. While children showed greater than expected growth in the measures used in the study, it may be that a longer time period between fall and spring assessments would have allowed for differences to emerge based on children's learning experiences in their early care and education settings.

Sampling. Several sampling issues limit the study findings. First, at the time of study recruitment, there were very few programs at the Bronze Circle of Quality; these programs could not be included in the study as originally planned, and therefore we do not know how aspects of quality at that level function. Second, there were few programs at the Silver Circle of Quality that entered through the accredited pathway so examination of pathway at the Silver level could not be conducted. Third, by system design, there were no PFA programs at the Licensed Circle. Fourth, the study included center-based programs only; ExceleRate enrolls family child care homes into the system, but because ExceleRate was not fully implemented when the validation study began, these programs were not yet part of the system and therefore could not be included in the study. We therefore do not know whether there are associations among Circles of Quality, quality measures, and children's outcomes for children in family child care homes.

Fifth, sampling of children was limited. This study included preschool-aged children only, although ExceleRate programs serve children from birth; resource limitations did not allow for inclusion of infants and toddlers in the study. Future evaluations of the system should focus on the extent to which ExceleRate captures aspects of quality that are important for infant and toddler development. Similarly, although the study attempted to recruit Spanish-speaking dual language learners, the sample size was very small. Future research should include a focus on addressing the extent to which ExceleRate Circles of Quality capture features of quality important for the learning and development of dual language learners.

Finally, related to sampling, the overall response rate was quite low. Of the programs contacted about the study and invited to participate, about 13% agreed to be in the study. Unmeasured differences between the programs that agreed and those that declined are likely to exist and have implications for the generalizability of the study results. Contextual factors likely contributed to challenges recruiting programs (see below).

Contextual factors. Two contextual factors were identified as impacting the study. First, data sharing agreements between the research team and state agencies for access to ExceleRate data were quite delayed, and not all data needed for the study were available to the study team. If the data sharing agreement had not been delayed, there might have been time to address the issue of insufficient data. As it was, we were not able to answer all questions, particularly the interesting questions related to weighting of standards and domains. For some questions, particularly those related to associations between domain scores and overall ratings, we were limited to using data on PFA programs only; results therefore represent only one sector and not the entire child care system. Second, the study occurred during a particularly challenging time for Illinois when the state was operating without a budget. We heard anecdotally that early education programs were affected by budget issues and may have been less likely to agree to participate in a study during this difficult period.

4.3. Conclusions and Recommendations for Consideration

Overall, the results of the validation study provide evidence that ExceleRate Circles of Quality meaningfully differentiate higher quality programs from lower quality programs. We found evidence of relations between ExceleRate Circles of Quality and independently observed quality across different quality measures, with no evidence that either program type or pathway was differentially related to the association between Circle and quality. The system seems to be operating as intended in that higher rated programs, regardless of program type or pathway, are of higher observed quality.

Regarding child outcomes, the study did not find associations between gains in children's assessment scores and Circles of Quality. This finding is not unexpected in a system that combines disparate measures into one overall rating. It is consistent with most other validation studies of statewide systems. A number of states are currently undertaking validation studies, and a few studies have been completed and released. Overall, states are finding limited, inconsistent, or nonexistent associations between ratings and outcomes. For example, researchers found that programs with higher ratings in Wisconsin's QRIS YoungStar did not produce larger gains in school readiness skills in a school year than programs with lower ratings (Magnuson & Lin, 2016). As other validation study reports are released, IL can learn from examination of these results. Although states vary in terms of their QRISs, and therefore findings from one state may not apply to another, it may be informative to examine whether patterns in the results could inform IL's efforts to revise ExceleRate Illinois.

Some research has suggested that greater attention should be paid to implementation of evidence-based curricula and learning activities that are structured in developmental sequences; such a focus might increase the ability of ExceleRate to better differentiate among programs that support children's developmental outcomes.

Illinois also might consider communication strategies for informing parents about scores. As noted above, a limitation of the study was the timing—validation activities occurred as the system was being rolled out, and parent information regarding awareness gathered by the study may reflect that limitation. Nevertheless, only 15% of parents in rated programs reported

that they considered the program's Circle of Quality prior to enrolling their child, and only 63% of directors reported that most parents in their program knew about their Circle of Quality rating. Directors reported that parents find out about ExceleRate ratings from a variety of sources, so there may be opportunity to increase parent access to information from sources like the ExceleRate website, with only 2% of directors identifying this as a source of information for parents about ExceleRate. Communicating about the Circles might include detailed information about domains and standards. In a block system like ExceleRate, programs might have higher ratings on individual domains, but their overall rating reflects their lowest domain score. Documenting and publishing the Circle of Quality levels for each of the four quality domains and their standards could be informative for parents.

Documenting Circle of Quality levels by domain and standards may also help programs to identify and focus on areas of strength and opportunities for quality improvement, which may lead to more targeted activities for achieving higher Circles of Quality. We found that 19% of the programs in the study sample not at the Gold Circle reported applying for a higher Circle of Quality, but fewer reported any structural barriers as reasons for not applying. It might be useful to consider reasons why so few programs are applying for higher Circles of Quality, and look for opportunities to support the 21% of programs that reported they are in the process of applying for a higher Circle of Quality rating. Follow-up interviews with directors of these programs might reveal more information about motivation to undertake the process of applying for a higher rating, barriers that they faced in the process, the type of help they received that was useful, and what supports they wished they had received

Overall, the results of the validation study suggest that ExceleRate Illinois is working as intended and can continue to serve as a framework for supporting IL's early childhood system. The intent of this study was to provide data to inform the continuous improvement of ExceleRate Illinois. It is hoped that the results will be used to refine the system and ensure that the QRIS is realizing its objectives for the children and families in Illinois.

Section 5. References

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American psychologist, 57(2),* 111.

Burchinal, M. R., Kainz, K., Cai, Y., Tout, K., Zaslow, M., Martinez-Beck, I., & Rathgeb, C. (2009). Early care and education quality and child outcomes. Research-to-Policy Research-to-Practice Brief: OPRE Research-to-Policy Brief #1. Washington, DC: U.S. Department of Health and Human Services, Administration on Children and Families, Office of Planning, Research and Evaluation, and Child Trends.

Burchinal, M. R., Roberts, J. E., Riggins, J. R., Zeisel, S. A., Neebe, E., & Bryant, D. (2000). Relating quality of center-based child care to early cognitive and language development longitudinally. Child Development, 71(2), 339-357. doi: 10.1111/1467-8624.00149

Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher–child interactions and instruction. Applied Developmental Science, 12(3), 140-153. doi: 10.1080/10888690802199418

Burchinal, M., Zaslow, M., & Tarullo, L. (2016). Quality thresholds, features, and dosage in early care and education: Secondary data analyses of child outcomes. *Monographs of the Society for Research in Child Development, 81*, 1-120

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the ability to remember what I said and to "do as I say, not as I do." *Developmental Psychobiology, 29*, 315-334. doi: 0012-1630/96/040315-20

Duncan, S. E., & DeAvilla, E. (1998). PreLAS. Montgomery, CA: McGraw-Hill.

Early, D.M., Bryant, D.M., Pianta, R.C., & Clifford, R.M. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten? *Early Childhood Research Quarterly, 21*, 174-195.

Fantuzzo, J., Perry, M. A., & McDermott, P. (2004). Preschool Approaches to Learning and Their Relationship to Other Relevant Classroom Competencies for Low-Income Children. *School Psychology Quarterly, 19(3),* 212.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network. (FMHI Publication No. 231).

Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., & Abner, K. (2013). An assessment of the validity of the ECERS-R with implications for measures of child care quality and relations to child development. Developmental psychology, 49(1), 146-160. doi: 10.1037/a0027899

Grissmer, D., & Kirby, S. N. (1997). Teacher turnover and teacher quality. Teachers College Record, 99(1), 45-56.

Harms, T., Clifford, R. M., & Cryer, D. (1998). The Early Childhood Environment Rating Scale (revised ed.). New York, NY: Teachers College.

Harms, T., Clifford, R. M., & Cryer, D. (2007). Infant/Toddler Environment Scale-Revised. New York: Teachers College Press.

Hindman, A. H., Skibbe, L. E., Miller, A., & Zimmerman, M. (2010). Ecological contexts and early learning: Contributions of child, family, and classroom factors during Head Start, to literacy and mathematics growth through first grade. Early Childhood Research Quarterly, 25(2), 235-250. doi: http://dx.doi.org/10.1016/j.ecresq.2009.11.003

Karoly, L. A. (2014). *Validation studies for early learning and care quality rating and improvement systems: A review of the literature*. Prepared for the Delaware Office of Early Learning. Rand Education.

Lower, J. K., & Cassidy, D. J. (2007). Child care work environments: The relationship with learning environments. Journal of Research in Childhood Education, 22(2), 189-204. doi: 10.1080/02568540709594621

Luria, A. R. (1966). Higher cortical functions in man. New York: Basic Books.

Magnuson, K., & Lin, Y.-C. (2016). *Validation of the QRIS YoungStar rating scale.* Retrieved from https://dcf.wisconsin.gov/files/youngstar/pdf/validationreport2.pdf

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., . . . Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. Child Development, 79(3), 732-749. doi: 10.1111/j.1467-8624.2008.01154.x

McDermott, P. A., Green, L. F., Francis, J. M., & Stott, D. H. (2000). *Preschool learning behaviors scale.* Philadelphia, PA: Edumetric and Clinical Science.

McDermott, P. A., Leigh, N. M., & Perry, M. A. (2002). Development and validation of the preschool learning behaviors scale. *Psychology in the Schools, 39(4)*, 353-365.

Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). The Bateria III Woodcock- Muñoz: Pruebas de aprovechamiento. Itasca, IL: Riverside.

Perkins, P., & Hightower, A. (2002). T-CRS 2.1 Teacher-Child Rating Scale: Examiner's Manual. Rochester, NY: Children's Institute.

Phillipsen, L. C., Burchinal, M. R., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of child care. Early Childhood Research Quarterly, 12(3), 281-303. doi: 10.1016/s0885-2006(97)90004-1

Pianta, R. C., Paro, K. M. L., & Hamre, B. K. (2008). The Classroom Assessment Scoring System. Manual, pre-K. Baltimore, MD: Brookes Publishing Co.

Ponitz, C. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly, 23,* 141–158. doi:10.1016/j.ecresq.2007.01.004.

Rubin, D. B. (1976). Inference and missing data. Biometrika, 63(3), 581-592. doi: 10.2307/2335739

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. New York, NY: John Wiley & Sons, Inc.

Sabol, T. J., & Pianta, R. C. (2014). Do standard measures of preschool quality used in statewide policy predict school readiness? Education Finance and Policy, 9(2), 116-164. doi: 10.1162/EDFP_a_00127

Sabol, T. J., Soliday Hong, S. L., Pianta, R. C., & Burchinal, M. R. (2013). Can rating pre-K programs predict children's learning? Science, 341(6148), 845-846. doi: 10.1126/science.1233517

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. Psychological Methods, 7(2), 147-177. doi: 10.1037/1082-989x.7.2.147

Schilder, D. (2017). Illinois Race to the Top Early Learning Challenge Evaluation: 2016 Evaluation Report. Boston, MA: BUILD Initiative.

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly, 22(2),* 173-187.

Talan, T.N., & Bloom, P.J. (2009). *Business Administration Scale.* New York, NY: Teachers College Press.

Talan, T.N., & Bloom, P.J. (2011). *Program Administration Scale-Second Edition*. New York, NY: Teachers College Press.

Thornburg, K. R., Mayfield, W. A., Hawks, J. S., & Fuger, K. L. (2009). The Missouri quality rating system school readiness study. Columbia, MO: Center for Family Policy & Research.

Tout, K., Cleveland, J., Li, W., Starr, R., Soli, M. & Bultinck, E. (2016). The Parent Aware Evaluation: Initial Validation Report. Minneapolis, MN: Child Trends.

Woodcock, R. W., & Muñoz-Sandoval, A. F. (1993). An IRT approach to cross-language test equating and interpretation. European Journal of Psychological Assessment, 9(3), 233-241.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Examiner's manual. Woodcock-Johnson III Tests of Achievement. Itasca, IL: Riverside Publishing.

Yazejian, N., Franco, X., Bryant, D., Morgan, J., Simpson, T., & Kuhn, L. (Forthcoming). North Carolina tiered quality rating and improvement system validation study final report. Chapel Hill, NC: Frank Porter Graham Child Development Institute, University of North Carolina at Chapel Hill.

Zellman, G. L., & Fiene, R. (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: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Zellman, G. L., & Perlman, M. (2008). Child-care quality rating and improvement systems in five pioneer states. Santa Monica, CA: RAND Corporation.

Zellman, G. L., Perlman, M., Le, V.-N., & Setodji, C. M. (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.