Research Brief OPRE 2016-110
December 2016

Measures Used in Quality Rating and Improvement Systems (QRIS) Validation Studies

Measures Used in Quality Rating and Improvement Systems (QRIS) Validation Studies

OPRE Research Brief #2016-110

December 2016

Submitted by: Kathryn Tout, Rebecca Starr, Jennifer Wenner, and Rowan Hilty

Child Trends

Submitted to: Ivelisse Martinez-Beck, PhD., Project Officer

Office of Planning, Research and Evaluation

Administration for Children and Families

U.S. Department of Health and Human Services

Contract Number: HHSP23320095631WC

Project Director: Kathryn Tout

Child Trends

7315 Wisconsin Avenue

Suite 1200 West

Bethesda, MD 20814

Suggested citation: Tout, K., Starr, R., Wenner, J. & Hilty, R. (2016). *Measures used in quality rating and improvement systems (QRIS) validation studies*. OPRE Research Brief #2016-110. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

This document was prepared to accompany other resources on evaluation of Quality Rating and Improvement Systems (QRIS) and other quality improvement initiatives developed by the Quality Initiatives Research and Evaluation Consortium (INQUIRE).

This report and other reports sponsored by the Office of Planning, Research and Evaluation are available at https://www.acf.hhs.gov/opre.

Overview

Investments in the development or revision of state and local Quality Rating and Improvement Systems (QRIS) for early care and education have increased in recent years, due in part to the Race to the Top – Early Learning Challenge (RTT-ELC) grants that were awarded to 20 states beginning in Fiscal Year 2011. The RTT-ELC grants included a requirement that states conduct evaluation activities to validate the rating process used by the QRIS to designate program quality levels and to assess the extent to which quality rating levels correlate with children's developmental outcomes.[1] Though a number of QRIS had conducted evaluation studies prior to 2011, the RTT-ELC grants resulted in multiple QRIS validation studies that were launched in 2012-2015. These validation studies assess the extent to which the QRIS rating process produces meaningfully distinct ratings of program quality and if quality rating levels correlate with children's developmental outcomes.

This research brief addresses the need for an aggregate picture of the methods, measures, and analytic strategies being used in QRIS validation studies by summarizing the measures that researchers are currently using or plan to use. To gather information on current QRIS validation study measures, Child Trends compiled data used to inform the Quality Initiatives Research and Evaluation Consortium (INQUIRE) validation workgroup, and reviewed published reports and research plans for RTT-ELC validation studies. A total of 19 QRIS validation studies were examined. All 19 of the validation studies are conducting analyses to assess how well items on the rating tool are working; 18 of the studies are assessing whether program quality ratings are different in meaningful ways; and, 16 studies are assessing whether quality ratings are related to measures of children's development. The Environment Rating Scales (ERS) and the Classroom Assessment Scoring System (CLASS) are the most commonly used observation tools in validation studies; in contrast, the set of child development measures used in validation studies varied across state. The brief also summarizes the structural quality indicators collected in the studies, covariates used in validation analyses, and some of the challenges that have been encountered in choosing a school readiness battery. The research brief was developed to inform discussions in the INQUIRE validation subgroup meetings but is being shared widely to facilitate a common understanding among QRIS stakeholders about QRIS validation studies.

[1] For details and further information about QRIS validation, please refer to a brief by Zellman and Fiene (2012) for a definition and overview, a literature review completed by Karoly (2014), a brief by Tout and Starr (2013) providing a template to support QRIS validation planning, and examples of validation results in four states (Lahti, Sabol, Starr, Langill & Tout, 2013).

Measures Used in Quality Rating and Improvement Systems (QRIS) Validation Studies

Purpose of the Brief

Investments in the development or revision of state and local Quality Rating and Improvement Systems (QRIS) for early care and education have increased in recent years, due in part to the Race to the Top – Early Learning Challenge (RTT-ELC) grants that were awarded to 20 states beginning in federal fiscal year 2011. The RTT-ELC grants required that states conduct evaluation activities to validate the rating process used by the QRIS to designate program quality levels, and that they assess the extent to which quality rating levels correlate with children's developmental outcomes.[2] Though a number of QRIS had conducted evaluation studies prior to 2011, the RTT-ELC grants resulted in multiple QRIS validation studies that were launched between 2012 and 2015. According to the QRIS Compendium, a comprehensive database of QRIS operating in the United States and its territories, 24 QRIS reported that they had either completed or were in the process of conducting a validation study in 2015.[3]

The purpose of this research brief is to summarize the measures that researchers are currently using or plan to use in a QRIS validation study. This need was identified by researchers involved in QRIS evaluations and validation studies. Researchers were invited to participate in the Quality Initiatives Research and Evaluation Consortium (INQUIRE), which serves as a learning community to support discussion and sharing of information about strategies used for, and challenges of, conducting QRIS research.[4] A subgroup of INQUIRE members meet regularly (typically monthly) to discuss research studies focused specifically on QRIS validation. In 2014, discussions in the validation subgroup highlighted the need for an aggregate picture of the methods, measures, and analytic strategies being used in QRIS validation studies.

[2] For details and further information about QRIS validation, please refer to a brief by Zellman and Fiene (2012) for a definition and overview, a literature review completed by Karoly (2014), a brief by Tout and Starr (2013) providing a template to support QRIS validation planning, and examples of validation results in four states (Lahti, Sabol, Starr, Langill & Tout, 2013).

[3] See http://qriscompendium.org/

[4] INQUIRE is supported by the Office of Planning, Research and Evaluation, Administration for Children and Families, through a contract with Child Trends.

The brief also provides information about the types of validation studies occurring, other indicators of quality collected, covariates used in validation analyses to control for selection factors and demographic characteristics, and some of the challenges that have been encountered in choosing a school readiness battery. The research brief was developed to inform discussions in the INQUIRE validation subgroup meetings, but it can also facilitate a common understanding among QRIS stakeholders about QRIS validation studies.

Overview of QRIS validation strategies

QRIS validation studies use different strategies to examine the extent to which the QRIS design (including the selected quality indicators) and implementation of the rating process produce accurate and meaningful ratings of programs (Karoly, 2014; Tout & Starr, 2013; Zellman & Fiene, 2012). Validation strategies can include activities such as a review of the empirical basis for the QRIS quality indicators used in the rating process, analysis of inter-correlations among the quality indicators, and assessment of procedures to support inter- and intra-rater reliability among staff that observe programs and/or review documentation gathered from programs.

This report focuses on the validation strategies that QRIS research teams are using to understand two issues raised in the RTT-ELC grant application. One set of strategies examines the extent to which the rating process produces ratings of program quality that are meaningfully distinct. For example, researchers may ask whether observed quality is different in programs with a 1-star rating compared to programs with a 4-star rating. Researchers may collect a measure of process quality[5] not used in the QRIS or only used at certain levels (for example, the Early Childhood Environment Rating Scale – Revised – ECERS-R; Harms, Clifford & Cryer, 2005) and examine the associations between the scores on the ECERS-R and program ratings. If expected associations are found with higher-rated programs showing higher ECERS-R scores and lower-rated programs showing lower ECERS-R scores, the analysis provides initial (though not complete) evidence of the validity of the ratings.

A second set of validation strategies examines the associations between QRIS ratings and measures of children's development, taking into account critical variables such as selection factors associated with children's participation in different early care and education programs and children's demographic characteristics. For example, researchers may collect direct assessments of children's language, math and cognitive skills, and teacher ratings of social/emotional skills in the fall and spring in the year before children transition to kindergarten. Children's gains in skills are expected to vary depending on the level of quality they experience.

Data suggesting that higher-quality programs are linked to larger gains in children's academic and social skills—after controlling for differences in children's characteristics and experiences that are also related to children's gains—would provide initial evidence of the validity of the QRIS. Yet there are multiple challenges and limitations of conducting child-level analyses in QRIS validation studies. For example, unobserved selection factors (such as parents' perceptions of early care and education quality; alignment with parenting practices; expectations for children's development; and motivations to seek different experiences for their children based on expectations) are difficult to control and may be predictors of children's development. In addition, measures are limited to assess the development of infants and toddlers and children who speak a language other than English. These and other factors should be considered when planning for validation strategies that involve child outcomes (Tout & Starr, 2013).

[5] Process quality is a broad term that includes features of the learning environment (such as materials, activities and health and safety provisions) and interactions (such as the extent to which teachers extend conversation and support children's language development).

As validation studies continue to be initiated across the country (particularly in response to RTT-ELC requirements), research teams make decisions about the measures to include and the analytic strategies to use. Because of the differences across QRIS, there has been no set of measures that are consistently used for either type of validation study (examination of program quality across rating levels and linkages between ratings and measures of children's development). This report provides an overview of measures and analytic strategies in use in QRIS validation studies, and is intended to support strategic planning and decision-making.

Method

The INQUIRE validation subgroup engaged in a variety of activities between October 2014 and August 2016 to compile the information included in this report. First, templates outlining measures used in each QRIS validation study were reviewed and filled in during monthly calls with researchers participating in the meetings.[6] Second, published reports and research plans for RTT-ELC validation studies were reviewed by researchers at Child Trends, and information was added to the templates. Third, an online survey tool was distributed to the monthly meeting participants to compile information intended to inform the monthly validation meeting discussions. Table 1 provides details about 22 QRIS validation studies known to be underway or recently completed. Nineteen of the current and/or planned validation studies are discussed in this brief. The remaining three studies are in the preliminary stages of planning or had insufficient information available.

Table 1. Validation studies included in the research brief.

State or county QRIS	Research team/Evaluation contractor	Study conducted for Race to the Top – Early Learning Challenge (yes or no)	Study status as of August 2016 and link to interim report if available
Arizona—Quality First*	Child Trends	Yes	In process
California—RTT-ELC QRIS*	American Institutes for Research	Yes	In process Interim report
Colorado—Colorado Shines*	Child Trends	Yes	In process
Delaware—Delaware Stars for Early Success*	RAND Corporation	Yes	In process Interim report
Georgia—Quality Rated*	Child Trends and Georgia State University	Yes	In process
Illinois—ExceleRate Illinois*	FPG Child Development Institute and American Institutes for Research	Yes	In process
Indiana—Paths to Quality Phase Two*	Purdue University	No	In process
Maryland—Maryland EXCELS*	Johns Hopkins University	Yes	In process
Massachusetts—Massachusetts Quality Rating and Improvement System*	UMass Donahue Institute and Wellesley College	Yes	In process

[6] Typically, 5 to 20 researchers participate in the monthly validation subgroup meetings. Researchers provided information about one or more validation studies that they direct.

Table 1 cont. Validation studies included in the research brief.

State or county QRIS	Research team/Evaluation contractor	Study conducted for Race to the Top – Early Learning Challenge (yes or no)	Study status as of August 2016 and link to interim report if available
Michigan—Great Start to Quality	McREL	Yes	In process
Minnesota—Parent Aware statewide expansion*	Child Trends	Yes	Brief Executive Summary Final report
Nevada—Nevada Silver State Stars QRIS*	University of Nevada	No	In process
New Jersey—Grow NJ Kids*	National Institute for Early Education Research	Yes	In process
New Mexico—FOCUS*	Child Trends	Yes	In process
North Carolina—Star Rated License System*	FPG Child Development Institute and Child Trends	Yes	In process
Ohio—Step Up To Quality	Compass Evaluation and Design	Yes	In process
Oregon—Oregon QRIS*	Oregon State University and Portland State University	Yes	In process
Pennsylvania—Pennsylvania Keystone STARS	University of Pennsylvania and the Consortium for Policy Research in Education	No	Report
Rhode Island—BrightStars*	Child Trends	Yes	In process
Vermont—Step Ahead Recognition System (STARS)*	Child Trends	Yes	In process
Washington—Early Achievers*	University of Washington	Yes	Executive Summary Report
Wisconsin—YoungStar*	University of Wisconsin	Yes	Executive Summary Report 1 Report 2

Source: Responses from INQUIRE Validation subgroup participants and Child Trends' review of existing documents
*Indicates a study with details described in this report.

Overview of QRIS validation studies

The following sections and accompanying tables provide a portrait of the QRIS validation studies that are currently underway or are planned for the future.

Goals of validation studies

QRIS validation studies vary in the overall purpose and objectives. Three objectives are identified in Table 2: (1) assess how the items in the rating tool are working (face validity, reliability, and internal consistency), (2) assess whether program quality rating levels are different in meaningful ways (concurrent validity), and (3) assess whether measures of children's development are related in expected ways to quality rating levels (predictive validity).

Table 2. Objectives of current and planned QRIS validation studies.

QRIS validation study	Assess how well items on the rating tool are working	Assess whether program quality ratings are different in meaningful ways	Assess whether measures of children's development are related to quality ratings
AZ—Quality First	X	X	X
CA—RTT-ELC QRIS	X	X	X
CO—Colorado Shines	X	X	
DE—Delaware Stars for Early Success	X	X	X
GA—Quality Rated	X	X	X
IL—ExceleRate Illinois	X	X	X
IN—Paths to Quality Phase Two	X	X	X
MD—Maryland EXCELS	X	X	X
MA—Massachusetts QRIS	X	X	X
MN—Parent Aware	X	X	X
NC—Star Rated License System	X	X	X
NJ—Grow NJ Kids	X	X	
NM—FOCUS	X	X	X
NV—Nevada Silver State Stars QRIS	X		X
OR—Oregon QRIS	X	X	X
RI—BrightStars	X	X	X
VT—Step Ahead Recognition System (STARS)	X	X	
WA—Early Achievers	X	X	X
WI—YoungStar	X	X	X
Total (19)	**19**	**18**	**16**

Source: Responses from INQUIRE Validation subgroup participants and Child Trends' review of existing documents, August 2016.

Nineteen QRIS validation studies are conducting analyses to understand how items on the rating tool are working. These analyses might include inter-correlations among the quality indicators to understand whether the items/indicators represent unique quality features or whether they are highly correlated with other items (which may indicate redundancy in the rating tool). In addition, distributions of scores on items might be examined to assess the difficulty of each item and the extent to which scoring reflects true quality differences or, alternatively, poorly articulated quality constructs. Eighteen QRIS validation studies include an objective to assess whether program quality ratings are different in meaningful ways. Sixteen QRIS validation studies aim to assess (in a current or future study) whether the QRIS ratings are related to child-level outcomes. Fifteen QRIS are conducting validation studies that address all three objectives.

Validation using program-level measures

An important question that is addressed in QRIS validation studies is whether program quality ratings are meaningful at each of the levels in a QRIS (Zellman and Fiene, 2012). Indeed, RTT-ELC grants require that states conduct a study to assess the question of whether rating levels differ in predicted ways. A well-functioning quality rating for an ECE program is expected to correlate with other measures designed to assess quality. For example, as rating levels increase, programs might be expected to score higher on established measures of observed quality. QRIS validation studies have typically used measures of global quality and adult-child interaction measures for this type of validation question (Karoly, 2014).

The most common measures used in validation research to examine whether quality ratings are meaningful at different rating levels are the Environment Rating Scales (ERS) and the Classroom Assessment Scoring System (CLASS).[7] The ERS are a group of observation measures designed to assess global quality by observing the environment, materials, routines, health and safety precautions, and teacher/caregiver-child interactions. In center-based programs, observers complete the Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford & Cryer, 1998) for preschool-aged classrooms, the Infant and Toddler Environment Rating Scale-Revised (ITERS-R; Harms, Cryer & Clifford, 1990) for infant/toddler classrooms, and the School-Age Care Environmental Rating Scale (SACERS; Harms, Jacobs, & White, 1996) for school-age classrooms. In family child care programs, observers use the Family Child Care Environment Rating Scale-Revised (FCCERS-R; Harms, Cryer & Clifford, 2007) to assess program quality.

Each environment rating scale has subscales (for example, a subscale on Space and Furnishings), and each subscale is comprised of items and indicators. Scores are computed for each subscale then averaged for a total program score. Research teams differ in how they use the ERS. Some researchers, for example, do not include the Parents and Staff subscale on the ECERS-R. Other differences in how items and indicators are scored on the ERS may also be apparent across research teams (and across QRIS implementation teams).

The Classroom Assessment and Scoring System-Pre-Kindergarten (CLASS - PreK; Pianta, La Paro & Hamre, 2008) is an observational tool used to assess the quality of emotional support and instruction in preschool classrooms. The CLASS is scored on three domains: Emotional Support (constructs such as the emotional connection between teachers and students, expressed negativity such as anger or hostility, and teacher sensitivity to students' concerns), Classroom Organization (behavior management, productivity, and instructional learning formats), and Instructional Support (concept development, how teachers provide feedback, and language modeling). Scores for each domain are based on a scale of 1 to 7. The CLASS-Toddler (Pianta, Hamre, & LaParo, 2010) version has two domains: Engaged Support for Learning and Emotional and Behavioral Support.

[7] Note that the ERS and the CLASS are also used within a QRIS to determine a program's rating level. Information about which tools are used in the QRIS ratings can be found at griscompendium.org.

Table 3 contains details about the tools that are being used in QRIS validation studies. Eighteen QRIS validation studies are currently and planning to use observation measures to understand whether and how quality ratings levels differ.

Table 3. Observation measures of quality used in QRIS validation studies.

QRIS validation study	ITERS-R	ECERS-R or ECERS-3	FCCERS-R	CLASS – Toddler	CLASS – Pre-K	CLASS – Combined	CIS	PQA
AZ—Quality First		X	X	X	X			
CA—RTT-ELC QRIS				X	X			X
CO—Colorado Shines	X	X	X					
DE—Delaware Stars for Early Success				X	X		X	X
GA—Quality Rated				X	X			
IL—ExceleRate Illinois	X	X			X			
IN—Paths to Quality Phase Two				X	X			
MD—Maryland EXCELS		X	X		X			
MN—Parent Aware		X[a]	X		X			
NC—Star Rated License System	X	X	X	X	X			
NJ—Grow NJ Kids	X*	X*		X*	X*			
NM—FOCUS	X*	X*			X*			
NV—Nevada Silver State Stars QRIS	X	X	X	X*	X*	X*		
OR—Oregon QRIS				X	X	X		
RI—BrightStars				X	X			
VT—Step Ahead Recognition System (STARS)	X	X	X					
WA—Early Achievers	X	X	X	X	X	X		

Table 3 cont. Observation measures of quality used in QRIS validation studies.

QRIS validation study	ITERS-R	ECERS–R or ECERS-3	FCCERS–R	CLASS – Toddler	CLASS – Pre-K	CLASS – Combined	CIS	PQA
WI—YoungStar		X	X					
Total programs currently using measure	6	10	9	9	12	2	1	2
Total programs considering using measure	2	2	0	2	3	1	0	0
Total (18)	**8**	**12**	**9**	**11**	**15**	**3**	**1**	**2**

Source: Responses from INQUIRE Validation subgroup participants and Child Trends' review of existing documents, August 2016.
*Indicates studies that are considering or planning to use respective measures/covariates.
[a] Indicates studies that also use the ECERS-E.

The most common measures currently used are the CLASS—Pre-K (12), ECERS-R or ECERS-3 (Harms, Clifford, & Cryer, 2014)(10), CLASS—Toddler (9), FCCERS-R) (9), and the ITERS-R (6). The CLASS-Infant is being used by 2 studies (OR and WA, not shown in Table 3). The Program Quality Assessment (PQA; Smith et al., 2005) and a combined version of the CLASS for both pre-K and toddlers are being used in two studies. Three additional tools are being used in one study each: the CLASS-Infant, the ECERS-E, and the Caregiver Interaction Scale (CIS; Arnett, 1989).

In addition to traditional measures of observed classroom quality and interactions, states are using newer measures to assess the language environment and the quality of child engagement and family engagement. The Oregon study currently uses the Individualized Classroom Assessment Scoring System (inCLASS; Downer et al., 2010) to observe student engagement and the Family Teacher Provider Relationship Questionnaire (FTPRQ; Kim et al., 2014) to measure family engagement. Maryland is collecting the Scale for Teacher's Assessment of Routines Engagement (STARE; McWilliam, 2000) to measure levels of children's engagement in the classroom. Washington uses the Language Environment Analysis (LENA; Xu, Yapanel, & Gray, 2009) to collect information about the language the child experiences. Washington also uses the Engagement in Classrooms Data Collection (ECDC; Taliano, Soderberg & Joseph, 2014) to track child engagement or on-task behaviors.

Three QRIS studies are still planning aspects of their work. The most common planned measure is the CLASS—Pre-K (3). The ITERS-R, ECERS-R or ECERS-3, and CLASS-toddler are each planned in two studies. The CLASS-Combined is being considered in Nevada.

Table 4 shows the use of measures designed to assess program management and administration. The Program Administration Scale (PAS; Talan & Bloom, 2011) is currently being used by two studies and the Business Administration Scale (BAS; Talan & Bloom, 2009) is currently used by one study.

Table 4. Measures of administration and management used in QRIS validation studies.

QRIS validation study	Program Administration Scale (centers)	Business Administration Scale (family child care)
IL—ExceleRate Illinois	X	
NC—Star Rated License System	X	X
Total	**2**	**1**

Source: Responses from INQUIRE Validation subgroup participants and Child Trends' review of existing documents, August 2016.

*Indicates studies that are considering or planning to use respective measures/covariates.

Indicators of structural quality

In addition to data collected on process quality (captured by the observational measures described in a previous section), validation studies typically include data collection about other indicators of quality. For example, data may be collected about the teacher/caregiver (e.g., educational qualifications, participation in recent training), the classroom (e.g., group size and ratio) or the program (e.g., accreditation status). These data may be used to validate the rating scale, to inform implementation, to check self-report data, as covariates in validation analyses, or for other uses.

As shown in Table 5, a total of 16 studies are currently using or plan to use structural quality indicators in their QRIS validation studies. Currently, structural quality indicators are incorporated into 13 validation studies. The most frequently included structural quality indicators are teacher/caregiver education level (13), teacher/caregiver certifications/credentials, and teacher's/caregiver's other professional development (11). Group size and adult-child ratio are each included in 10 studies. Salary and/or benefits are included in six validation studies and teacher/caregiver turnover is included in five studies.

Table 5. Structural quality indicators used in QRIS validation studies.

QRIS validation study	Teacher/ caregiver education level	Teacher/ caregiver certifications/ credentials	Teacher's/ caregiver's other professional development	Teacher/ caregiver turnover	Teacher/ caregiver salary and/ or benefits	Group size	Adult-child ratio
CA—RTT-ELC QRIS	X		X				
CO—Colorado Shines	X	X	X			X	X
GA—Quality Rated	X	X	X			X	X
IL—ExceleRate Illinois	X	X	X		X	X	X
IN—Paths to Quality Phase Two	X	X	X			X	X
MD—Maryland EXCELS	X	X	X	X		X	X
MA—Massachusetts QRIS	X	X	X		X	X	X

Table 5 cont. Structural quality indicators used in QRIS validation studies.

QRIS validation study	Teacher/ caregiver education level	Teacher/ caregiver certifications/ credentials	Teacher's/ caregiver's other professional development	Teacher/ caregiver turnover	Teacher/ caregiver salary and/ or benefits	Group size	Adult-child ratio
MN—Parent Aware	X	X		X	X		
NC—Star Rated License System	X	X	X		X	X	X
NJ—Grow NJ Kids	X*		X*				
NM—FOCUS	X*	X*	X*	X*	X*		
NV—Nevada Silver State Stars QRIS	X	X	X			X	X
OR—Oregon QRIS	X	X	X	X	X		
RI—BrightStars	X			X		X	X
VT—Step Ahead Recognition System (STARS)	X*	X*	X*			X*	X*
WI—Youngstar	X	X	X	X	X	X	X
Total programs currently using measure	13	11	11	5	6	10	10
Total programs considering using measure	3	2	3	1	1	1	1
Total (16)	**16**	**13**	**14**	**6**	**7**	**11**	**11**

Source: Responses from INQUIRE Validation subgroup participants and Child Trends' review of existing documents, August 2016.

*Indicates studies that are considering or planning to use respective measures/covariates.

Three studies plan to incorporate structural quality indicators in their QRIS research. Each of these three plan to include teacher/caregiver education level and teacher's/caregiver's other professional development. Two plan to include teacher/caregiver certifications/credentials. One study each may include teacher/caregiver turnover, teacher/caregiver salary and/or benefits, group size, and adult/child ratio.

Validation using child development measures

Another key validation question is the extent to which measures of children's developmental outcomes are correlated in expected ways with QRIS program ratings (Zellman & Fiene, 2012). Based on the existing early care and education literature highlighting the role of ECE program quality in supporting children's development (particularly for low-income children), the expectation in a QRIS is that children served by higher-rated programs will make greater gains on measures of child development (or demonstrate overall academic and social skills at a higher level) than children served in lower-rated programs (taking into account selection factors and child and family characteristics that are also associated with outcomes).

RTT-ELC grants require states to conduct a validation analysis (or to show how the analysis is being planned for the future) that includes measures of children's development. Researchers typically choose a battery of child assessments to measure multiple domains including executive function, language and literacy, math concepts, and social/emotional development. In states with a significant proportion of Spanish-speaking children, Spanish versions of the assessment are administered. Children attending QRIS-rated

programs may be assessed at one time point or at two time points (e.g., fall and spring) for a measure of developmental gain. The following sections describe the child development measures being used in current QRIS validation studies.

Executive function measures

Some QRIS validation studies include measures of executive function in the child assessment battery. Executive function includes cognitive constructs such as attention, self-regulation, rule-following, and short-term memory. These constructs are related to general cognitive abilities and are increasingly included in assessments of school readiness.

As shown in Table 6, 10 studies use or plan to use measures of executive function. The Pencil-Tap Test (Diamond & Taylor, 1996) is included in five validation studies and may be incorporated into one more. Six studies are currently using Head-Toes-Knees-Shoulders (McClelland et al., 2007).

Table 6. Measures of executive function used in QRIS validation studies.

QRIS validation study	The Pencil-Tap test	Head-Toes-Knees-Shoulders
CA—RTT-ELC QRIS	X	
DE—Delaware Stars for Early Success		X
GA—Quality Rated		X
IL—ExceleRate Illinois	X	X
MN—Parent Aware	X	
NC—Star Rated License System	X	X
NM—FOCUS	X*	
RI—BrightStars	X	
WA—Early Achievers		X
WI—YoungStar		X
Total programs currently using measure	5	6
Total programs considering Using measure	1	0
Total (10)	**6**	**6**

Source: Responses from INQUIRE Validation subgroup participants and Child Trends' review of existing documents, August 2016.

*Indicates studies that are considering or planning to use respective measures/covariates.

Language screener

A language screener is sometimes used as part of a child assessment battery to ensure that a child is proficient enough in the language in which the tools are administered to be able to continue testing. Five validations studies are using the Pre-Language Assessment Scale (Pre-LAS; Duncan & De Avila, 1998) for this purpose, as shown in Table 7.

Table 7. Language screener measure (Pre-Language Assessment Scale) used in QRIS validation studies.

QRIS validation study	Pre-LAS
CA—RTT-ELC QRIS	X
DE—Delaware Stars for Early Success	X
IL—ExceleRate Illinois	X
MN—Parent Aware	X
WA—Early Achievers	X
Total (5)	**5**

Source: Responses from INQUIRE Validation subgroup participants and Child Trends' review of existing documents, August 2016.

*Indicates studies that are considering or planning to use respective measures/covariates.

Social/emotional development measures

Measures of social/emotional development are an integral part of child assessment batteries designed to test school readiness. Table 8 shows that eleven studies are or plan to incorporate social/emotional development measures. The Social Competence and Behavior Evaluation (SCBE, SCBE-30; LaFreniere & Dumas, 1995) and the Preschool Learning and Behavior Scale (PLBS; McDermott, Leigh, & Perry, 2002) are each being used in four validation studies. Three QRIS validation studies are using the Devereux Early Childhood Assessment (DECA; Mackrain, LeBuffe,& Powell, 2007), while one study plans to use it in the future. The Teacher-Child Rating Scale (TCRS; Hightower et al., 1986) is being used in three studies. The Brief Infant Toddler Social Emotional Assessment (BITSEA; Briggs-Gowen & Carter, 2006) is being used in Indiana's validation and the Social Skills Improvement System-Rating Scale (SSIS; Gresham & Elliot, 2008) is used in Wisconsin's study. The Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2000) for preschoolers and infants/toddlers is being used in Washington's study.

Table 8. Social/emotional development measures used in QRIS validation studies.

QRIS Validation Study	SCBE/ SCBE-30	PLBS	DECA	TCRS	CBCL
DE—Delaware Stars for Early Success			X		
GA—Quality Rated			X	X	
IN—Paths to Quality Phase Two	X				
MA—Massachusetts Quality Rating and Improvement System		X	X		
MN—Parent Aware	X	X			
NC—Star Rated License System				X	
NM—FOCUS	X*	X*			
NV—Nevada Silver State Stars QRIS			X*		
RI—BrightStars	X	X			
WA—Early Achievers				X	X

Table 8 cont. Social/emotional development measures used in QRIS validation studies.

QRIS Validation Study	SCBE/ SCBE-30	PLBS	DECA	TCRS	CBCL
WI—YoungStar	X	X			
Total programs currently using measure	4	4	3	3	1
Total programs considering using measure	1	1	1	0	0
Total (11)	**5**	**5**	**4**	**3**	**1**

Source: Responses from INQUIRE Validation subgroup participants and Child Trends' review of existing documents, August 2016.

*Indicates studies that are considering or planning to use respective measures/covariates.

Language and literacy measures

Language and literacy measures are a central component of a school readiness child-assessment battery and are included in thirteen current and planned studies. Presently, 10 validation studies are using at least one measure of language or literacy (see Table 9). Four studies are using the Peabody Picture Vocabulary Test-Fourth Edition (PPVT-4; Dunn & Dunn, 2007), a measure of receptive vocabulary. Letter-Word Identification, a subtest of the Woodcock Johnson (WJ-III: Woodcock, McGrew & Mather, 2001; WJ-IV: Woodcock & Munoz-Sandoval, 1993) is being used in eight studies. Three studies are using the Picture-Vocabulary subtest of the Woodcock-Johnson. Two studies are using Test of Preschool Early Literacy (TOPEL; Lonigan, Wagner, Torgeson, & Rashotte, 2007). Georgia is using two teacher reports of language development for infants and toddlers: the LENA Snapshot (Language Environment Analysis System; Gilkerson & Richards, 2008) and the MacArthur-Bates Communicative Inventories (CDI; Fenson et al. 2006). Other language and literacy measures are currently used in one study are:

- Indicators of individual Growth and Development for Infants and Toddlers (IGDI; Center of Early Education and Development, 1998) (MN)
- Story and Print Concepts (Mason & Stewart, 1989)(CA)
- Academic Rating Scale – Language and Literacy (Rock & Pollack, 2002) (IL)
- Early Writing Assessment (EWA) (WA)

Three studies are planning to use various language and literacy measures in their validation studies including: Letter-Word Identification, Woodcock Johnson (3), PPVT-4 (2), and IGDI Picture Naming (1).

Table 9. Language and literacy measures used in QRIS validation studies.

QRIS validation study	PPVT–4	Letter Word Identification (WJ)	Picture Vocabulary (WJ)	IGDI Picture Naming	TOPEL
CA—RTT-ELC QRIS		X[a]			
DE—Delaware Stars for Early Success	X	X[a]			
GA—Quality Rated		X*[b]		X*	
IL—ExceleRate Illinois		X[a]	X		
IN—Paths to Quality Phase Two	X				
MA—Massachusetts Quality Rating and Improvement System	X	X[a]			
MN—Parent Aware				X	X
NC—Star Rated License System		X[a c]	X		
NM—FOCUS	X*[d]	X*[a c]			
NV—Nevada Silver State Stars QRIS	X*	X*[a]			
RI—BrightStars		X[a]	X		
WA—Early Achievers	X[d]	X[ac]			
WI—Youngstar		X[a]			X
Total programs currently using measure	4	8	3	1	2
Total programs considering using measure	2	3	0	1	0
Total (13)	**6**	**11**	**3**	**2**	**2**

Source: Responses from INQUIRE Validation subgroup participants and Child Trends' review of existing documents, August 2016.

*Indicates studies that are considering or planning to use respective measures/covariates.

[a] Indicates studies that use the WJ-III.

[b] Indicates studies that use the WJ-IV.

[c] Indicates studies that use the Woodcock-Munoz for Spanish-speaking children.

[d] Indicates studies that use the TVIP for Spanish-speaking children (Dunn et al. 1986).

Math concept measures

Early math concepts are typically included in school readiness batteries. Table 10 displays the 11 current and planned studies that include one math measure, the Woodcock Johnson-III or Woodcock Johnson-IV subtest, Applied Problems (Woodcock, McGrew & Mather, 2001; Woodcock & Munoz-Sandoval, 1993). This is being used in eight studies currently and three other studies plan to use Applied Problems in their future validation work. In addition, Illinois used the teacher-reported Academic Rating Scale – Mathematical Thinking (Rock & Pollack, 2002). Washington's study used a different measure of early math – the Tools for Early Assessment in Math (TEAM; developed by Clements and Sarama). Washington also uses the Lens on Science (LENS) to assess science learning.

Table 10. Math concept measures used in QRIS validation studies.

QRIS validation study	Applied Problems (WJ)
CA—RTT-ELC QRIS	X[a]
DE—Delaware Stars for Early Success	X[a]
GA—Quality Rated	X*[b]
IL—ExceleRate Illinois	X[a]
MA—Massachusetts Quality Rating and Improvement System	X[a]
MN—Parent Aware	X[a]
NC—Star Rated License System	X[a c]
NM—FOCUS	X*[c]
NV—Nevada Silver State Stars QRIS	X*
RI—BrightStars	X[a]
WI—YoungStar	X[a]
Total programs currently using measure	8
Total programs considering using measure	3
Total (11)	**11**

Source: Responses from INQUIRE Validation subgroup participants and Child Trends' review of existing documents, August 2016.

*Indicates studies that are considering or planning to use respective measures/covariates.

[a] Indicates studies that use the WJ-III.

[b] Indicates studies that use the WJ-IV.

[c] Indicates studies that use the Woodcock-Munoz for Spanish-speaking children.

Physical development

One validation study (Minnesota) is using Body Mass Index (BMI) as a physical development measure. The QRIS validation in Illinois plans to incorporate a parent's report of child's health status and a teacher report of the child's general health.

Measures covering multiple domains

Some assessments are designed to measure multiple domains of school readiness. For example, the Bracken School Readiness Assessment (Bracken, 1998) provides a school readiness score that encompasses several concepts such as recognition of colors, letters, and numbers.

Table 11 shows the use of measures covering multiple domains of school readiness in six current and planned validation studies. Three studies include the Bracken School Readiness Assessment and it is planned in another study. Two studies use the Woodcock Johnson—III Tests of Achievement, which is being considered in one other validation as well. The validation studies in Indiana and Washington are using the Mullen Scales of Early Learning (Mullen, 1989) for infants and toddlers.

Table 11. Measures covering multiple domains used in QRIS validation studies.

QRIS validation study	Woodcock Johnson—III Tests of Achievement (WJ–III)	Bracken	Mullen Scales of Early Learning
CA—RTT-ELC QRIS	X		
IN—Paths to Quality Phase Two		X	X
MN—Parent Aware	X	X	
NM—FOCUS	X*	X*	
WA—Early Achievers			X
WI—Youngstar		X	
Total programs currently using measure	2	3	2
Total programs considering using measure	1	1	0
Total (6)	**3**	**4**	**2**

Source: Responses from INQUIRE Validation subgroup participants and Child Trends' review of existing documents, August 2016.

*Indicates studies that are considering or planning to use respective measures/covariates.

Use of child assessment measures

Relative to program-level measures, there is a greater variety of child-level measures used across QRIS validation studies. Language and literacy measures are commonly included in child assessment batteries, but the selected measures vary across studies (most commonly, Letter-Word Identification from WJ-III or WJ-IV (11) and the PPVT-4 (6)). Several other language/literacy measures are used in a couple of studies.

The second most frequently-included domain in child assessment batteries is math concepts. When studies include a math assessment, most use Applied Problems (WJ-III or WJ-IV). Ten studies include or plan to include a measure of executive functioning: The Pencil Tap Test (6 and Head-Toes-Knees-Shoulders (6). Eleven studies are using or considering a measure of social/emotional development in their child

assessment battery (six different measures). Six studies incorporate assessments that assess multiple domains and five studies include a language screener (the Pre-LAS) and. Measures of physical development in child assessment batteries are not commonly used in current QRIS validation studies.

Given the variety of measures used for child assessment in validation studies, it is helpful to explore common considerations and challenges in developing child assessment batteries. Researchers participating in the INQUIRE validation subgroup have discussed a variety of issues. First, it is challenging to select a comprehensive assessment battery that can be administered efficiently. Researchers aim to develop a short battery that respects young children's attention span and minimizes their time away from the ongoing activities in their ECE program.

Similarly, researchers do not want to place undue burden on teachers and caregivers by administering assessment batteries that are too long. The group discussed strategies for structuring the battery (for example, including a more active assessment in the middle of the battery to keep the child's attention).

Overall, researchers discussed the importance of selecting measures that meet the following criteria:

- assess key domains of children's school readiness,[8]
- do not overlap with other measures in the battery,
- align with measures used in other studies
- have been normed for children from low-income households and from different racial and ethnic backgrounds,
- have evidence of reliability and validity,
- are engaging and fun for children to complete,
- cover relevant child age ranges,
- are available in English and Spanish, and
- are easy for assessors to be trained on and to administer

The INQUIRE validation researchers relayed the importance of checking in with other researchers about their protocols and implementing lessons learned from studies they and others have conducted.

In addition to considerations related to development of the child assessment battery, researchers also discussed the analytic challenges of examining child outcomes in early care and education programs. For example, data collection and analytic models must take into account selection factors related to participation in early care and education programs with low and high ratings and the child and family characteristics associated with developmental outcomes. Factors such as children's attendance are particularly important to include in analytic models but may be challenging to collect reliably. Researchers also acknowledged the need to collect measures at two points in time to allow for a measure of growth rather than a point in time score. These and other design and analytic issues will be discussed in more depth in a future brief.

[8] Researchers in the validation subgroup have noted the challenge of selecting measures of school readiness in validation studies when the state does not have an established definition of school readiness.

Covariates used in QRIS validation analyses

In QRIS validation research, data are often collected on variables to be used as covariates in the analysis models. For example, when researchers want to look at the relationship between QRIS level and child development, they need to control for other factors that predict child development, such as family income or parent education level. When these and other factors are controlled for, a clearer understanding of the relationship between QRIS and program- or child-level measures can be obtained. A variety of program-level, classroom-level, and child/family-level covariates are being used in QRIS validation studies.

Program level covariates

Table 12 displays the 14 QRIS validation studies that are using or considering program-level covariates including program type, program size, and program location or urbanicity.

Table 12. Program-level covariates used in QRIS validation studies.

QRIS validation study	Center-based vs. family child care home	Other program types	Program size	Program location or urbanicity
CA—RTT-ELC QRIS	X	X	X	
CO—Colorado Shines	X	X	X	X
DE—Delaware Stars for Early Success	X	X	X	X
GA—Quality Rated	X		X	X
IL—ExceleRate Illinois		X	X	X
IN—Paths to Quality Phase Two	X	X		
MD—Maryland EXCELS	X*	X*	X*	X*
MA—Massachusetts QRIS			X*	X*
MN—Minnesota Parent Aware	X	X	X	X
NC—Star Rated License System	X	X	X	X
NJ—Grow NJ Kids		X*		X*
NM—FOCUS	X*			
OR—Oregon QRIS	X	X	X	X
VT—Step Ahead Recognition System (STARS)	X	X	X	X
Total programs currently using measure	9	9	9	8
Total programs considering using measure	2	2	2	3
Total (14)	**11**	**11**	**11**	**11**

Source: Responses from INQUIRE Validation subgroup participants and Child Trends' review of existing documents, August 2016.

*Indicates studies that are considering or planning to use respective measures/covariates.

Classroom and teacher/caregiver-level covariates

Classroom and teacher/caregiver-level variables, such as ages served and demographic characteristics are used and planned as covariates in twelve validation studies. Currently, 11 studies are using these covariates (see Table 13). The most common classroom and teacher/caregiver characteristics used are age group served (10) and teacher/caregiver length of time in field/experience (9). Classroom/group composition and primary language spoken (by teacher/caregiver and/or in program) are each included in six studies. Teacher/caregiver race/ethnicity is being used in five studies. Teacher/caregiver age, teacher/caregiver gender, teacher's/caregiver's feelings of support by administration/leaders and teacher's/caregiver's professional attitudes/plan for career in field are included in three studies each. Other classroom and teacher/caregiver-level covariates are being used in one state each:

- provider Stage of Change Scale[9](IN),
- director's/administrators' experience/credentials/time at the site (MD), and
- topics of training, mentoring, and professional development in the past year (OR).

In addition, one study is planning to use classroom/teacher/caregiver-level covariates including age group served, classroom/group composition, and teacher/caregiver length of time in field/experience.

Child-level covariates

Child (or family)-level covariates such as child gender, race/ethnicity, household income, and parent education are also included in 15 current and planned QRIS validation analyses (see Table 14). Presently, 11 studies include child-level covariates. Gender is the most common covariate (included in eleven studies). Nine studies each use pretest scores, dual language learner status, family income, and parent education. Race/ethnicity is included in eight studies. Disability status and hours per week in classroom/group are included in seven studies and length of attendance is included in six studies. Three studies include the number of people in the household. Two studies control for if the child also attends another program. These five child level covariates are only used by one state each:

- parental marital status (MN),
- parental immigrant status (MN),
- neighborhood covariates (i.e. average household income, education) (VA),
- age of entry to non-parental care (MD), and
- child temperament (OR).

Four studies plan to use child-level covariates. The most common covariates planned are gender, race/ethnicity, and dual language learner status (each planned in 4 studies), hours per week in the classroom/group (3) and whether the child attends another program (3). Pretest scores, disability status, length of attendance in the classroom/group, family income, and parent education may be incorporated into two studies each. Maryland's validation may include the number of people in household.

[9] The Stage of Change Scale assesses how child care providers perceive their current level of engagement in change of their practices. Peterson, S.M., Baker, A., & Weber, M (March, 2010). *Stage of Change Scale for Early Education and Care 2.0 Professional Manual.* Children's Institute. Rochester, NY.

Table 13. Classroom level covariates used in QRIS validation studies.

QRIS validation study	Age group served	Classroom/ group composition	Teacher/ caregiver age	Teacher/ caregiver race/ ethnicity	Teacher/ caregiver gender	Teacher/ caregiver length of time in field/ experience	Teacher's/ caregiver's feelings of support by administration/ leaders	Teacher's/ caregiver's professional attitudes/ plan for career in field	Primary language spoken (by teacher/ caregiver and/or in program)
CA—RTT-ELC QRIS	X					X			X
CO—Colorado Shines	X	X		X		X			
DE—Delaware Stars for Early Success	X								
GA—Quality Rated	X[a]					X[a]	X[a]	X[a]	X[a]
IL—ExceleRate Illinois	X	X		X	X	X			X
IN—Paths to Quality Phase Two	X		X			X			
MD—Maryland EXCELS	X	X	X	X	X	X	X	X	X
MA—Massachusetts QRIS	X*	X*				X*			
NC—Star Rated License System	X	X		X		X			X
OR—Oregon QRIS	X	X	X	X	X	X	X	X	X
RI—Bright Stars						X			

Table 13 cont. Classroom level covariates used in QRIS validation studies.

QRIS validation study	Age group served	Classroom/ group composition	Teacher/ caregiver age	Teacher/ caregiver race/ ethnicity	Teacher/ caregiver gender	Teacher/ caregiver length of time in field/ experience	Teacher's/ caregiver's feelings of support by administration/ leaders	Teacher's/ caregiver's professional attitudes/ plan for career in field	Primary language spoken (by teacher/ caregiver and/or in program)
VT—Step Ahead Recognition System (STARS)	X	X							
Total programs currently using measure	10	6	3	5	3	9	3	3	6
Total programs considering using measure	1	1	0	0	0	1	0	0	0
Total (12)	**11**	**7**	**3**	**5**	**3**	**10**	**3**	**3**	**6**

Source: Responses from INQUIRE Validation subgroup participants and Child Trends' review of existing documents, August 2016.

*Indicates studies that are considering or planning to use respective measures/covariates.

[a] Indicates a study that is collecting the measure but has not yet determined whether it will be covaried in statistical models.

Table 14. Child-level covariates used in QRIS validation studies.

QRIS validation study	Pretest scores	Gender	Race/ ethnicity	Dual language learner status	Disability status	Length of attendance in classroom/ group	Hours per week in classroom/ group	Does child also attend another program?	Family income	Number of people in household	Parent education
CA—RTT-ELC QRIS	X	X		X	X						
DE—Delaware Stars for Early Success	X	X	X	X	X	X	X		X		X
GA—Quality Rated	X	X[a]	X[a]	X[a]	X[a]	X[a]	X[a]		X[a]		
IL—ExceleRate Illinois	X	X	X	X	X	X	X		X		X
IN—Paths to Quality Phase Two	X	X			X	X	X	X	X	X	X
MD—Maryland EXCELS		X*	X*	X*	X*	X*	X*	X*	X*	X*	X*
MA—Massachusetts QRIS		X*	X*	X*	X*		X*	X*			
MN—Mineesota Parent Aware		X	X	X							X
NC—Star Rated License System	X	X	X	X					X		X
NM—FOCUS	X*	X*	X*	X*					X*		X*
NV—Nevada Silver State Stars QRIS	X*	X*	X*			X*	X*	X*			
OR—Oregon QRIS		X	X	X	X	X	X		X		X
RI—Bright Stars	X	X	X	X*			X		X	X	X
WA—Early Achievers	X	X		X					X		X
WI—YoungStar	X	X	X	X	X	X	X	X	X	X	X

Table 14 cont. Child-level covariates used in QRIS validation studies.

QRIS validation study	Pretest scores	Gender	Race/ ethnicity	Dual language learner status	Disability status	Length of attendance in classroom/ group	Hours per week in classroom/ group	Does child also attend another program?	Family income	Number of people in household	Parent education
Total programs currently using measure	9	11	8	9	7	6	7	2	9	3	9
Total programs considering using measure	2	4	4	4	2	2	3	3	2	1	2
Total (15)	**11**	**15**	**12**	**13**	**9**	**8**	**10**	**5**	**11**	**4**	**11**

Source: Responses from INQUIRE Validation subgroup participants and Child Trends' review of existing documents, August 2016.

*Indicates studies that are considering or planning to use respective measures/covariates.

[a] Indicates a study that is collecting the measure but has not yet determined whether it will be covaried in statistical models.

Conclusion

The purpose of this research brief is to examine the measures that are being used in QRIS validation studies that are underway or being planned. By examining the patterns across studies, researchers can inform the selection of measures for their own validation studies.

Currently, 16 studies are using program-level measures to validate the QRIS rating scale and eleven studies include child-level measures. The most commonly used program-level measures of quality are the CLASS-PreK (12 studies) and the ECERS-R or ECERS-3 (10 studies). Other versions of the Environment Rating Scales (FCCERS-R and ITERS-R) are also used frequently in validation. The CLASS-toddler version is being used by nine studies. A similar pattern was noted for those studies still in the planning process.

Over half of the studies include child-level outcomes, but there is a variety of measures used in validation studies. Language and literacy measures are the most frequently used assessments, particularly Letter-Word Identification from the Woodcock Johnson-III or WJ-IV (eight studies) and the PPVT (four studies). Several other language/literacy measures are used in other validation studies. There was less variety in measures used for assessing math concepts, with Applied Problems (WJ-III or WJ-IV) currently being used in eight of the nine studies assessing math.

Currently, nine studies use one or more measures of social/emotional development, five studies are using a language screener, and nine studies use a measure of executive function. Measures of physical development are being used in only two QRIS validation studies. The primary challenges researchers encounter in creating a child assessment battery is maximizing coverage of "school readiness" constructs while minimizing time burden on children.

Validation researchers also include other measures of quality in their studies. Data on structural quality, such as teacher/caregiver characteristics or adult/child ratio, can be used in several ways. For example, data can be used to validate any self-reported data or to use as covariates in analyses. The most commonly used structural quality indicators are teacher/caregiver characteristics (i.e., education, credentials, and other professional development) followed by program characteristics (i.e., group size and ratio).

Covariates are included in validation studies to control for factors that may predict the outcomes of interest. The most commonly used program-level covariates are program type and teacher/caregiver characteristics. At the child/family level, family income, child gender, race/ethnicity, pre-test scores, and parent education are commonly used covariates.

As new validation studies are launched, it will be important to catalogue the selection of measures and identify any new trends, particularly in the selection of child assessment measures. And, when results of current validation studies begin to be published, it will be important to conduct a similar scan across studies to identify patterns in the results. This synthesis of results will inform not only decisions about future research but also decisions about QRIS design and implementation.

References

Achenbach, T.M., & Rescorla, L.A. (2000). Manual for the ASEBA Preschool Forms & Profiles. Burlington, VT: University of Vermont Department of Psychiatry.

Arnett, J. (1989). Caregivers in day care centers: Does training matter? Journal of Applied Developmental Psychology, 10, 541–552.

Bracken, B. (1998). Bracken Basic Concept Scale Revised: Examiner's Manual. London: The Psychological Corporation.

Briggs-Gowan, M. J., & Carter, A. S. (2006). The Brief Infant-Toddler Social and Emotional Assessment (BITSEA). San Antonio, TX: Psychological Corporation, Harcourt Assessment.

De Avila, E. A., & Duncan, S. E. (1998). Pre-School Language Assessment Scale (Pre-LAS 2000—Spanish). Monterey, CA: CTB/McGraw-Hill.

Diamond, A. & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to do as I say, not as I do. *Developmental Psychobiology, 29,*315-334.

Downer, J.T., Booren, L.M., Lima, O.K., Luckner, A.E., Pianta, R.C.. The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. Early Childhood Research Quarterly.2010;25(1):1–16

Dunn, L.M., & Dunn, D.M. (2007). The Peabody Picture Vocabulary Test, Fourth Edition. Bloomington, MN: NCS Pearson, Inc.

Dunn, L. M., Padilla, E. R., Lugo, D. E., Dunn, L. M., & American Guidance Service (Firm). (1986). *TVIP : Test de Vocabulario en Imagenes Peabody: Adaptacion Hispanoamericana = Peabody Picture Vocabulary Test [Revised] : Hispanic-American adaptation*. Circle Pines, Minn: American Guidance Service.

Early Childhood Research Institute on Measuring Growth and Development. (1998). Research and Development of Individual Growth and Development Indicators for Children between Birth and Age Eight. Technical Report No. 4. Minneapolis, MN: Center for Early Education and Development, University of Minnesota. www.myigdis.com.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2006). The MacArthur–Bates Communicative Development Inventories (2nd ed.). Baltimore, MD: Brookes.

Gilkerson, J., & Richards, J.A. (2008b). *The LENA™ developmental snapshot (LENA Foundation Technical Report LTR-07-2)*. Retrieved from LENA Foundation: http://www.lenafoundation.org/TechReport.aspx/Snapshot/LTR-07-2.

Gresham, F.M. & Elliot, S.N. (2008). Social Skills Improvement System Rating Scales manual. Pearson, Inc.:MN

Harms, T., Clifford, R.M., Cryer, D. (1998). Early Childhood Environment Rating Scale, Revised Edition. Teachers College Press; New York.

Harms, T., Clifford, R.M., and Cryer, D. (1990, 2003). Infant Toddler Environment Rating Scale, Revised Edition. New York: Teachers College Press, Columbia University.

Harms, T., Clifford, R. M., Cryer, D. (2014). Early Childhood Environment Rating Scale, Third Edition. Teachers College Press, Columbia University.

Harms, T., Jacobs, E.V., and White, D.R. (1996). School-Age Care Environment Rating Scale. New York: Teachers College Press, Columbia University.

Harms, T., Cryer, D., & Clifford, R. M. (2007). Family Child Care Environment Rating Scale- Revised Edition. New York, NY: Teachers College Press.

Hightower, A. D., Work, W. C., Cowen, E. L., Lotyczewski, B. S., Spinell, A. P., Guare, J. C., & Rohrbeck, C. A. (1986). The Teacher-Child Rating Scale: A brief objective measure of elementary children's school problem behaviors and competencies. School Psychology Review, 15(3), 393-409.

Howes, C., & Stewart, P. (1987). Child's play with adults, toys, and peers: An examination of family and child care influences. Developmental Psychology, 23, 423-430.

Invernizzi, M., Sullivan, A., Meier, J. (2001) Phonological awareness literacy screening: Preschool (PALS-Pre-K). Charlottesville, VA: University of Virginia.

Ireton, H. (1992). Child Development Inventory Manual. Minneapolis, Minnesota: Behavior Science Systems, INC.

Karoly, Lynn A. (2014). Validation Studies for Early Learning and Care Quality Rating and Improvement Systems: A Review of the Literature. Santa Monica, CA: RAND Corporation. http://www.rand.org/pubs/working_papers/WR1051.

Kim, K., Porter, T., Atkinson, V., Rui, N., Ramos, M., Brown, E., Guzman, L., Forry, N., and Nord, C. (2014). Family and Provider/Teacher Relationship Quality Measures: User's Manual. OPRE Report 2014-65. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

LaFreniere, P.J. & Dumas, J.E. (1995). Social Competence and Behavior Evaluation, preschool edition. Los Angeles: Western Psychological Services. (60 pages).

Lahti, M., Sabol, T., Starr, R., Langill, C., & Tout, K. (2013). Validation of Quality Rating and Improvement Systems (QRIS): Examples from Four States, Research-to-Policy, Research-to-Practice Brief OPRE 2013-036. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. (2007). The Test of Preschool Early Literacy. Austin, TX: Pro-Ed.

Mackrain, M., LeBuffe, P., & Powell, G. (2007). Devereux early childhood assessment for infants and toddlers. Lewisville, NC: Kaplan Early Learning Company.

Mason, J.M. & Stewart, J. (1989). Story and Print Concepts. The CAP Early Childhood Diagnostic Instrument (prepublication edition).

McClelland, M.M., Cameron, C. E., Connor, C. M., Farris, C.L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*, 947-959

McDermott, P. A., Leigh N. M., & Perry, M.A. (2002). Development and validation of Preschool Learning Behaviors Scale. *Psychology in the Schools, 39*, 353-365

Mullen, E. M. (1989). Infant Mullen scales of early learning. Cranston, RI: T.O.T.A.L. Child.

O'Brien, J. (2003, July). The national reporting system and Head Start children with disabilities. Head Start Bulletin, 76, 16-17, 33.

Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). Classroom Assessment Scoring System (CLASS) Pre-K Version. Baltimore, MD: Brookes Publishing.

Pianta, R. C., Hamre, B. K., & LaParo, K. M. (2010). Classroom Assessment Scoring System (CLASS) Toddler Version. Baltimore, MD: Brookes Publishing.

Reid, D., Hresko, W., & Hammill, D. (2001). Test of Early Reading Ability-Third Edition. Pro-ED, Texas: 2001.

Rock, D. A., & Pollack, J. M. (2002). Early Childhood Longitudinal Study: Kindergarten class of 1998–99 (ECLS-K). Psychometric report for kindergarten through first grade (NCES No. 2002-05). Washington, DC: National Center for Education Statistics, Department of Education.

Sherman, Elisabeth M. S. and Brooks, Brian L.(2010) 'Behavior Rating Inventory of Executive Function-Preschool Version (BRIEF-P): Test Review and Clinical Guidelines for Use', Child Neuropsychology, 16: 5, 503 — 519, First published on: 04 May 2010 (iFirst)

Smith, Charles, Akiva, Tom, Sugar, Samantha A., Lo, Yun-Jia, Frank, K. A., Peck, Steve C., & Cortina, Kai. (2012). Continuous quality improvement in afterschool settings: Impact findings from the Youth Program Quality Intervention study. Ypsilanti, MI: Forum for Youth Investment.

Sylva, K., Siraj-Blatchford, I., & Taggart, B. Early Childhood Environment Rating Scale, Extended Edition. Teachers College Press; New York: 2010.

Talan, Teri N., & Bloom, Paula J. (2011). *Program Administration Scale: Measuring early childhood leadership and management* (2nd ed.). New York: Teachers College Press.

The Build Initiative & Child Trends. (2014). A Catalog and Comparison of Quality Rating and Improvement Systems (QRIS) [Data System]. Retrieved from http://qriscompendium.org/ on January 27, 2015.

Talan, T.N. & Jorde Bloom, P. Business Administration Scale for Family Child Care. Teachers College Press; New York: 2009.

Tout, K. & Starr, R. (2013). Key Elements of a QRIS Validation Plan: Guidance and Planning Template. OPRE 2013-11. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III Tests of Cognitive Abilities. Itasca, IL: Riverside.

Woodcock, R. W., & Muñoz-Sandoval, A. F. (1993). *Woodcock-Muñoz Language Survey, English Form*. Riverside.