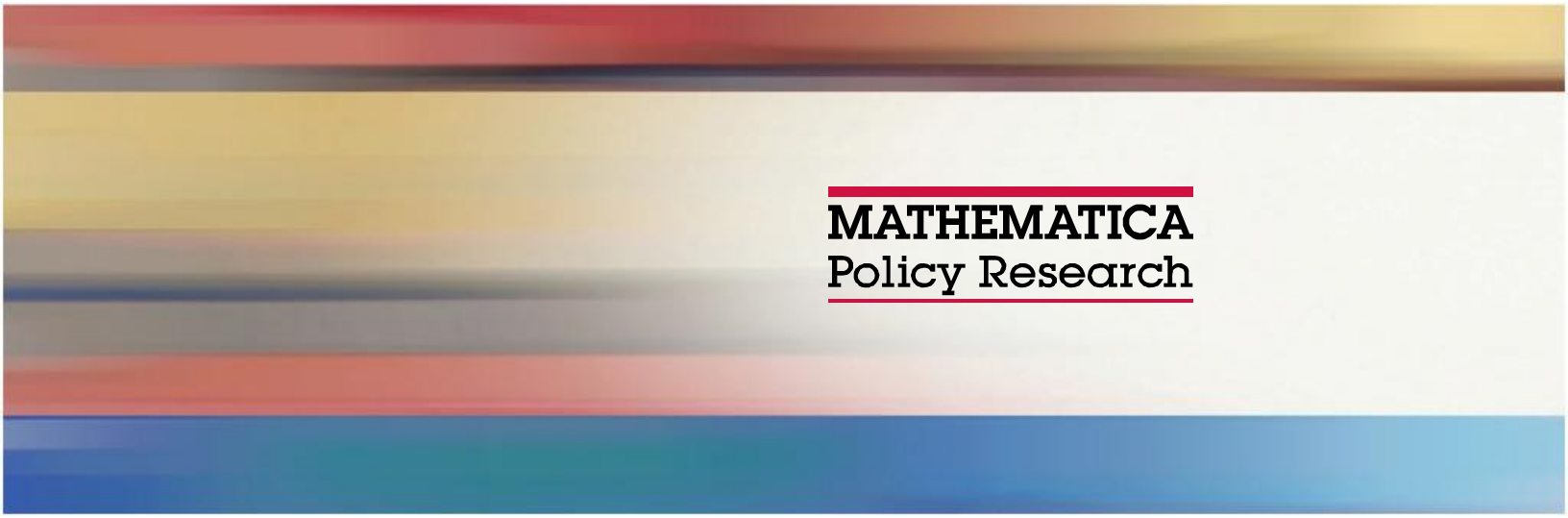
# Race to the Top- Early Learning Challenge: Descriptive Study of Tiered Quality Rating and Improvement Systems in Nine Round 1 States: OMB Forms Clearance Package, Part B

April 21, 2015



MATHEMATICA
Policy Research

Contract Number:
ED-IES-10-C-0077

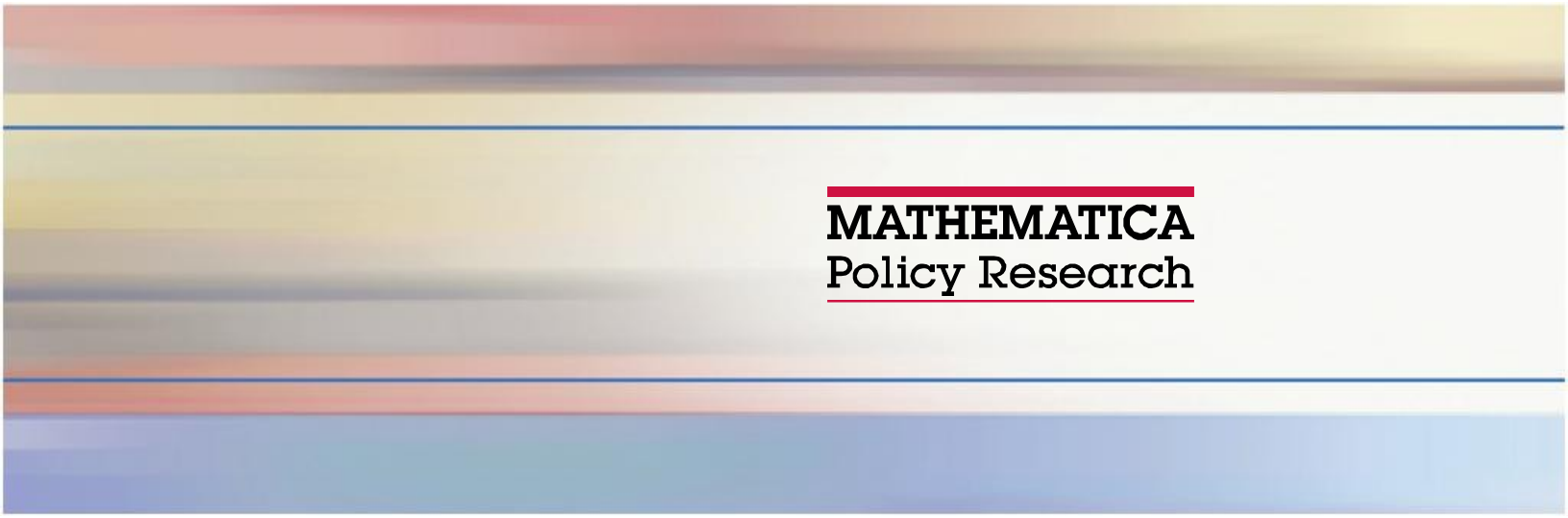
Mathematica Reference Number:
06844.143

Submitted to:
National Center for Education Evaluation
555 New Jersey Ave., NW, Suite 500k
Washington, DC 20208
Project Officer: Tracy Rimdzius

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005
Project Director: Susanne James-Burdumy, PhD

# Race to the Top- Early Learning Challenge: Descriptive Study of Tiered Quality Rating and Improvement Systems in Nine Round 1 States: OMB Forms Clearance Package, Part B

April 7, 2015



## CONTENTS

## TABLES

## PART B. SUPPORTING STATEMENT FOR PAPERWORK REDUCTION ACT SUBMISSION

This supporting statement requests clearance to recruit and collect data from approximately 14 school districts to support the Study of Race to the Top–Early Learning Challenge Tiered Quality Rating and Improvement Systems (RTT-ELC TQRIS). Through TQRIS, states use a set of standards to evaluate the quality of early learning and development programs and promote program improvement. The RTT-ELC application for the first round of grants required states to develop and adopt a TQRIS that, according to the U.S. Department of Education (ED), "is based on a statewide set of tiered program standards that include: (1) early learning and development standards, (2) a comprehensive assessment system, (3) early childhood educator qualifications, (4) family engagement strategies, (5) health promotion practices, and (6) effective data practices; is clear and has standards that are measurable, meaningfully differentiate program quality levels, and reflect high expectations of program excellence commensurate with nationally recognized standards that lead to improved learning outcomes for children; and is linked to the state licensing system for early learning and development programs" (2011). The study of RTT-ELC TQRIS will provide important information on the structure and characteristics of TQRIS and whether and how the TQRIS ratings relate to child outcomes. This information is intended to inform improvements to states' TQRIS. The Institute of Education Sciences (IES) at ED contracted with Mathematica Policy Research to conduct this study, which is being carried out as a component of the Impact Evaluation of Race to the Top and School Improvement Grants.[1]

The RTT-ELC TQRIS study comprises two distinct phases. The first study phase is a descriptive analysis of the the TQRIS in the nine Round 1 RTT-ELC states. In each state, we will examine the structure of the TQRIS, including components and indicators included in states' TQRIS and how they are combined to generate overall quality ratings.[2] We will also explore the rigor and consistency in the

[1] The Office of Management and Budget (OMB) has already approved recruitment and data collection activities for the Impact Evaluation of Race to the Top and School Improvement Grants. Recruitment activities were approved under OMB# 1850-0884 on June 27, 2011. Data collection activities were approved under OMB# 1850-0884 on April 4, 2012.

[2] TQRIS standards are commonly grouped into categories or components. Examples of components of TQRIS standards include health and safety, curriculum and learning, environment, staff qualifications, family engagement, and administration and management. In TQRIS, programs are typically assigned component-level ratings that are then combined to determine an overall quality rating. Each component is composed of indicators, which are the specific criteria used to assess a program.

measurement process (both in defining the indicators and collecting the rating information), the influences on program entry into and movement within each system (such as incentives), and key elements that administrators continue to refine and why. This phase will include the nine round 1 RTT-ELC states and rely on data gathered from their grant applications, reviews of public and state-provided documents, and targeted discussions with RTT-ELC grant coordinators (or their designees) in these nine states.[3] *(Clearance is not sought for these state-level data collection activities, as they require only targeted interactions with fewer than 10 respondents.)* In gathering information on the structure and characteristics of TQRIS, we will also assess the feasibility of examining the relationship between characteristics of center-based early care and education (ECE) programs (as measured by TQRIS ratings) and children's kindergarten readiness outcomes. If this proves feasible, the second study phase will examine the relationship between TQRIS ratings and child outcome measures in two to three of the Round 1 RTT-ELC states where the conditions exist to make such analysis possible. For this phase of the study, data on TQRIS ratings, component-level ratings, and indicator-level ratings will come from state TQRIS databases. Data on kindergarten entry assessments will also be sought from databases administered at the state-level. In the event that the kindergarten data needed for the analysis are not available from state databases, it may be necessary to reach out to selected districts in the RTT-ELC states to collect such data. If this step proves necessary, the contractor estimates (based on study power requirements and prior recruiting experiences) needing to reach out to up to 42 districts in order to ultimately recruit 14 districts from which to collect outcomes data. Because more than nine entities may need to be included in such recruitment and data collection activities, this supporting statement seeks clearance for the materials and activities that would support the study's *district* recruitment efforts and *district*-level data collection *if these activities are needed*. This document includes drafts of the district recruitment letter (Appendix A), the RTT-ELC TQRIS study information sheet (Appendix B), district memorandum of understanding (Appendix C), district recruitment call protocol (Appendix D), and district data request form (Appendix E).

[3] Documents will include TQRIS websites, and documents received from the RTT-ELC states (such as program guides, TQRIS standards, and descriptions of TQRIS data systems). We will also review documents shared by the RTT-ELC program office (including state evaluation plans and annual performance reports) and request information pertaining to the nine states from ongoing research efforts, such as the QRIS online compendium.

## B. Collection of Information Requiring Statistical Methods

### 1. Respondent universe and sampling methods

The RTT-ELC grants program is a joint effort by the U.S. Departments of Education (ED) and Health and Human Services (HHS) to strengthen the quality of early learning and development programs, and to increase access to high quality programs for all children, particularly low-income and disadvantaged children. The states that were awarded RTT-ELC grants were required to commit to integrating resources and aligning policies across the various state agencies that administer public funds related to early learning and development. RTT-ELC seeks to improve early learning and development programs by supporting reforms in five areas: (1) integrating state early care and education systems; (2) increasing access to high-quality, accountable programs; (3) promoting early learning and development outcomes through statewide standards and comprehensive assessment systems; (4) developing and supporting the early childhood workforce; and (5) building early learning data systems to measure child outcomes and progress.

ED and HHS have awarded RTT-ELC grants through three rounds of competition. Notices for the first competition—which was open to all states, the District of Columbia, and Puerto Rico—were posted in August 2011. In that first round of competition, 35 states, the District of Columbia, and Puerto Rico applied for RTT-ELC grants.[4] Round 1 awards were made in December 2011 to nine states —California, Delaware, Maryland, Massachusetts, Minnesota, North Carolina, Ohio, Rhode Island, and Washington. Across these states, a total of $500 million dollars was awarded, with grants ranging from $45 million to $70 million. A second round of RTT-ELC grants was announced in September 2012. Because more modest funding was available, the second round invited the next five highest scoring finalists from Round 1—Colorado, Illinois, New Mexico, Oregon, and Wisconsin—to modify their 2011 applications and create plans that could be funded with up to 50 percent of the amounts originally requested. All five invited states applied and, in December 2012, were awarded RTT-ELC grants in amounts ranging from $21 million to $35 million. A third round of competition was announced in April 2013. Awards ranging from $36 million to $51 million were made in December 2013 to six new states—Georgia, Kentucky, Michigan, New Jersey, Pennsylvania, and Vermont. In addition, supplementary grants were provided to six states from rounds 1 and 2 (California, Colorado, Illinois, New

[4] Alabama, Alaska, Idaho, Indiana, Louisiana, Montana, New Hampshire, North Dakota, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, and Wyoming did not apply for RTT-ELC grants in the first round.

Mexico, Oregon and Wisconsin) that had previously only received 50 percent of the funds they requested. This study focuses on the nine states awarded RTT-ELC grants in Round 1.

A key aspect of the RTT-ELC program is the development and implementation of TQRIS under the second reform area noted above (that is, increasing access to high quality, accountable programs). TQRIS aim to improve child care program quality and child outcomes by promulgating rigorous standards, rating providers based on those standards, and publicizing the ratings of individual providers (Zellman et al. 2011). The usefulness of TQRIS ratings thus depends on their ability to reliably measure the components that comprise the overall rating and construct from them an overall rating that is a valid indicator of program differences with regard to child outcomes (Cizek 2007, as cited in Zellman and Feine 2012).

Despite growing interest in the use of TQRIS to identify high quality early learning programs, evidence is limited as to whether TQRIS ratings distinguish quality and the extent to which the TQRIS ratings are associated with child outcomes. To our knowledge, five studies have examined the relationship between child outcomes and TQRIS ratings (Zellman et al. 2008; Thornburg et al. 2009; Elicker et al. 2011; Tout et al. 2011; Sabol and Pianta 2012, as cited in Lahti et al., 2013), and all examine TQRIS developed prior to the RTT-ELC program. Only two of these five studies—Thornburg et al. (2009) and Sabol and Pianta (2012)—find evidence of a correlation between TQRIS ratings and outcomes. Thornburg and colleagues (2009) found that children attending programs with 3-, 4- or 5-star ratings from Missouri's TQRIS showed significantly greater gains during the preschool year in social skills and behavioral outcomes relative to children attending programs with 1- or 2-star ratings. They did not find statistically significant differences in children's vocabulary, early literacy, or math skills. Sabol and Pianta (2012) found that children who attended programs with 3- or 4-star ratings from Virginia's TQRIS exhibited significantly larger increases in literacy skills over the prekindergarten year compared with children attending programs with 2-star ratings. To learn about TQRIS in states that have received RTT-ELC grants and to inform ongoing development of states' TQRIS, ED has contracted with Mathematica Policy Research to conduct this study.

The study of RTT-ELC TQRIS is designed to answer the following research questions:

1. How are TQRIS ratings defined, collected, and generated in the nine Round 1 RTT-ELC states and how might the structure and implementation of the rating system influence the reliability and validity of the ratings?
2. Is it feasible to examine the relationship between TQRIS ratings and children's kindergarten readiness outcomes in selected RTT-ELC states?
3. What is the relationship between TQRIS ratings and kindergarten readiness outcomes for programs that serve preschool-age children in selected RTT-ELC states?

The study will involve two samples: one for research questions 1 and 2 and one for research question 3.

### a. Select states and districts

**Sample for descriptive analysis and feasibility assessment.** The study's first phase will consist of all nine states that are RTT-ELC Round 1 states. The descriptive analysis and feasibility assessment will rely on three data sources: (1) states' RTT-ELC applications, (2) publicly available and other existing documents that contain information on each state's TQRIS and kindergarten entry assessments, and (3) interviews with state representatives who have in-depth knowledge of the state's TQRIS, kindergarten entry assessments, and data systems for the TQRIS and KEA.

#### *(1) RTT-ELC Applications*

In February through April 2013, we reviewed RTT-ELC Applications for Initial Funding for the nine Round 1 RTT-ELC states. From the RTT-ELC applications, we gathered information about the structure of each state's TQRIS, plans for expanding program participation in TQRIS, type and use of kindergarten entry assessments, and existing data systems.

#### *(2) Publicly available and other existing documents*

From September through December 2014, we collected and conducted a targeted review of documents pertaining to the structure of TQRIS and characteristics of kindergarten entry assessments in the nine Round 1 RTT-ELC states as well as the availability of TQRIS and kindergarten entry assessment data. Documents were collected from three sources: (1) the RTT-ELC program office and ongoing research efforts, (2) state websites and (3) state officials. Most documents were available on state websites. To access documents that were not publicly available but offer important details about the TQRIS (such as revised

standards that were not yet publicly available) and kindergarten entry assessments, we requested copies from state officials during initial study engagement conversations in September and October 2014.

***(3) Interviews with state representatives***

From December 2014 through March 2015, we conducted telephone interviews with state representatives from the nine Round 1 RTT-ELC states to confirm and clarify the information obtained from documents and, as needed, gathered information that could not be obtained from the document reviews. During initial study engagement conversations, the team identified the appropriate representatives to interview. The team interviewed between two to four administrators in each state to cover the four primary topics: (1) TQRIS rating structure, (2) TQRIS data, (3) kindergarten entry assessments, and (4) kindergarten entry assessment data. For example, in some states, the team interviewed the state TQRIS administrator to confirm or gather details about the quality rating components and overall rating structure and separately interviewed a researcher or data manager to understand details about data coverage and accessibility that the TQRIS administrator could not address. In other states, the TQRIS administrator was able to address topics related to both the TQRIS rating structure and the characteristics of the data. No topic or related series of questions was asked of more than one respondent in each of the nine states. The team developed and used tailored protocols for each respondent, not only to reduce burden, but also to ensure that the same (or similarly worded) information was not collected from more than nine individuals – thereby allowing the information collected from State Representatives to be exempt from the Paperwork Reduction Act (PRA).

**Sample for analyses of the relationship between TQRIS ratings and child outcome Measures**. We estimate that the sample for the analysis of TQRIS ratings and measures of children's school readiness will include two to three of the nine RTT-ELC Round 1 states or districts within those states. These states and/or districts will be purposefully selected based on the availability, quality, depth, and coverage of the data required to conduct these analyses, which include: (1) TQRIS data, including overall quality level ratings and ratings on components and indicators that comprise the overall rating; (2) kindergarten entry assessment data; (3) a common identifier (at the student and/or program level) in the TQRIS and kindergarten entry assessment data sets; and (4) data on child demographics and other characteristics.

***(1) TQRIS data***

States included in this study component will need to provide TQRIS ratings for participating center-based early care and education (ECE) programs. In particular, we will require TQRIS ratings for a sufficient number of center-based ECE programs for which we can also link kindergarten entry assessment data. These data linkages might be more feasible for state-funded center-based ECE programs, because children who attend these programs may be assigned a unique state or district identifier as a preschooler that is later linked to kindergarten assessment data. We will also explore other ways to link the necessary data. If it is possible to link the necessary data for other types of center-based ECE programs – beyond school-based or other state-funded programs – such as licensed private child care centers—we will explore including these programs in our analysis.

In general, states with larger numbers of participating center-based ECE programs will be selected over states with smaller numbers. Furthermore, states with higher variance of TQRIS ratings among center-based ECE programs will be selected over states with lower variance.[5] Both of these factors affect statistical power and will therefore be considered when selecting states for the analysis. During our conversations with state officials, we will therefore request updated information and ask follow-up questions to determine how many of the programs currently participating in the states' TQRIS are center-based ECE programs whose data could be linked to kindergarten entry assessment data and the distribution of ratings among such programs.

We will obtain data on center-based ECE programs' TQRIS ratings in 2014–2015 from state TQRIS databases. In addition to overall TQRIS ratings, if available, we will obtain data on component-level ratings. Data on component-level ratings are necessary for examining associations between the components of TQRIS ratings and child outcomes.

### *(2) Kindergarten entry assessment data*

The second requirement for state inclusion in this study component is the use of a standard kindergarten assessment across the state (or within targeted districts) and the availability of kindergarten entry assessment data for children who were enrolled in center-based ECE programs that participate in the state's

[5] Some TQRIS have special considerations for assigning quality ratings to state-funded preschool programs. For example, state-funded preschool programs in Minnesota can obtain the highest rating of 4 through an "accelerated pathway." In North Carolina, state-funded preschool programs are required by law to have a rating of 4 or 5 (the top two ratings). Because special considerations can severely restrict the variance in TQRIS ratings, we will explore these issues carefully for each state.

TQRIS. Ideally, kindergarten entry assessments will measure several of the five essential domains of school readiness referenced in the RTT-ELC applications: (1) language-literacy, (2) cognition and general knowledge (including early mathematics and early scientific development), (3) approaches toward learning, (4) physical well-being and motor development (including adaptive skills), and (5) social and emotional development. For each child with kindergarten entry assessment data, we will also need to determine whether he/she attended a TQRIS-participating center-based ECE program and, if so, identify the program. We describe potential strategies for obtaining this information in the next section. During our interviews with state representatives, we will investigate what percentage of students with kindergarten entry assessment data were previously enrolled in a center-based ECE program that participated in the TQRIS. In general, states with larger percentages of students meeting these criteria will be selected over states with smaller percentages.

We plan to use kindergarten entry assessment data from tests administered in the fall of 2015. If kindergarten entry assessment and TQRIS data from the prior year (that is, fall 2014 kindergarten entry assessments and 2013-2014 TQRIS ratings) are available, *and* if the kindergarten entry assessments and TQRIS standards used in that year were the *same* as those used in the next year (fall 2015 kindergarten entry assessments and 2014–2015 TQRIS ratings), we will request data for both years to increase the sample for analysis.

Ideally, kindergarten entry assessment data will come from a centralized source at the state level. This approach would allow us to obtain data for all center-based ECE programs that participated in the TQRIS and whose students went on to attend kindergarten in the state. It would provide a larger sample size and, hence, a higher level of statistical power to detect a statistically significant relationship between TQRIS ratings and kindergarten entry assessment measures.

If we cannot obtain the necessary kindergarten entry assessment data from a centralized source at the state level, we will attempt to collect such data from individual school districts. However, collecting kindergarten entry assessment data from districts would necessitate that we limit data collection to districts that contain the highest proportion of TQRIS-participating center-based ECE programs (and variability in TQRIS ratings among those programs) in the state.[6] If collecting data from districts proves necessary, we anticipate obtaining kindergarten entry assessment data from a targeted sample of 14 districts with 200 *state-funded*

[6] This approach to selecting districts will help maximize statistical power to detect differences between the highest- and lowest-quality programs.

preschool programs.[7,8] Based on recruiting rates from similar prior IES studies, we further anticipate that we would need to reach out to about 42 districts to ultimately recruit 14 into the study. We will provide states and/or districts with a form that specifies the data needed for the study (Appendix E).

### *(3) A common identifier (at the student or program level) in the TQRIS and kindergarten entry assessment data sets*

The third requirement for state inclusion in the analysis of TQRIS ratings and child outcomes is that there exists in the TQRIS and kindergarten entry assessment data sets a common identifier that will allow us to link program-level TQRIS data to child-level kindergarten entry assessment data. Specifically, we will need to be able to match a kindergartener's assessment data with the overall TQRIS rating for the center-based ECE program that he/she attended in the previous year. We expect that these data linkages will be more feasible for children who attended state-funded preschool programs, because the children might be assigned a unique state or district identifier that is linked to their kindergarten data. However, there may be other ways to link center-based ECE program TQRIS ratings with kindergarten assessment data. For example, the kindergarten entry assessment data set may include an ID for the center-based ECE program that a student attended, and it may be possible to use that program ID to obtain the program's rating from the TQRIS data set. It may also be possible to obtain center-based ECE enrollment rosters from states or districts that contain student IDs that match to the student IDs used in the state or district kindergarten entry assessment databases, so that TQRIS data can be linked with child outcome data.

### *(4) Data on child demographics and other characteristics*

[7] According to the Common Core of Data, on average, the number of schools serving preschool-aged children in districts serving preschool-aged children is 14, so we are assuming that we will need to obtain data from 14 districts in order to include roughly 200 state-funded preschools in our sample (14 = 200/14). We may be able to include other typed of center-based ECE programs in our analysis if we can link their TQRIS ratings with kindergarten entry assessment data. However, if we need to obtain data from districts, we will target districts based on the number of *state-funded* preschools, since it is more likely that we will be able to link TQRIS ratings to kindergarten entry assessment data for children who attended such programs.

[8] It is possible that only some of the two to three selected states will be able to provide the needed data and that we will obtain the data from districts in the other state(s). The sample sizes shown here assume that none of the selected states can provide the needed data and that we would need to collect all the necessary data from districts. If some states can provide data, then the sample for these analyses would be larger than 200 programs.

A final factor that will inform the selection of states for inclusion in the second study component is the availability of data on child demographics and other characteristics to be included in the regression analysis. These data will be useful to (a) control for factors other than differences in center-based ECE program quality that might be correlated with child outcomes and (b) conduct subgroup analyses (focusing, for example, on students from low-income households, minorities, and English language learners). Thus, we will select states and/or districts with the most comprehensive, high-quality data on child characteristics.

We will explore the availability of data on gender, birth year, socioeconomic status, race and ethnicity, English language learner status, free or reduced-price lunch status, special education status, and maternal education. We will also inquire about the availability of data on children's length of enrollment in their center-based ECE programs, which would make it possible to control for differences in program dosage in our analysis.

### b. Recruiting states and districts

The recruitment process for the RTT-ELC TQRIS study will occur in two stages. First, we will reach out to each of the nine Round 1 RTT-ELC states to inform them of the study; participation in the study is a requirement of the RTT-ELC grant. The second stage, district recruitment, will be conducted only if needed. In this section, we first briefly describe the state recruitment process for context. The majority of this section discusses district recruitment, which is the focus of this supporting statement.

**State engagement.** In September and October 2014, we contacted each of the nine Round 1 RTT-ELC states to introduce the study, identify respondents for telephone interviews, and request documents pertaining to the structure of TQRIS and characteristics of kindergarten entry assessments in the nine Round 1 RTT-ELC states as well as the availability of TQRIS and kindergarten entry assessment data. We sent introductory packages to the RTT-ELC grant coordinators in each state that, which will included:

- A **state engagement letter** that explained the study's importance and provided an overview of the study.
- An **information sheet** that provided a nontechnical description of the study. This document included a summary of study activities, the study's research questions, and a timeline for study activities.

A recruiter assigned to the state followed up with the state grant coordinator to confirm receipt of the mailing and to conduct a brief (approximately half-hour) telephone discussion with the appropriate agency staff person. Recruiters used a standardized protocol as a guide during these telephone conversations, which presented background information on the study, requested specific documents, and gathered contact information for use in scheduling telephone interviews. All nine state representatives agreed to participate in phase 1 of the study during these engagement calls and provided the appropriate contacts with whom to schedule the interviews.

**District recruitment.** If we learn through the interviews with state representatives that we cannot obtain the necessary kindergarten entry assessment data from a centralized source at the state level, we will attempt to collect such data from individual school districts. If collecting data from districts proves necessary, we anticipate obtaining kindergarten entry assessment data from a targeted sample of 14 districts with 200 *state-funded* preschool programs. Based on recruiting rates from similar previous IES studies, we further anticipate that we would need to reach out to 42 districts to ultimately recruit 14 into the study. For all targeted districts, the contractor will send introductory packages to the district superintendent, which will include:

- A **district recruitment letter** (Appendix A) that explains the study's importance, provides an overview of the study, including estimated participation time and confidentiality information, and indicates that a member of the study team will call to provide more details.
- An **information sheet** (Appendix B) that provides a nontechnical description of the study. This document includes a summary and timeline of study activities, the study's research questions, time and obligation to participate, and confidentiality information.
- A **district memorandum of understanding** (Appendix C) that describes the agreed-upon roles and responsibilities of the study team and of participating districts.

Recruiters will make follow-up calls to the district contacts to introduce the study and arrange for a time to further discuss the study. Recruiting communications with each district will take place by phone and email. Recruitment discussions will last approximately 1.5 hours and will be guided by a standardized recruitment protocol (Appendix D). The protocol will guide recruiters in describing the study design and planned data collection activities, as well as gathering information on the characteristics of center-based ECE

programs in the district and the availability of district-level kindergarten entry assessment data. Based on the information gathered during these discussions with district officials, the study team will identify 14 districts to participate in the second study phase. We estimate that these recruiting efforts will take place in fall 2015 (as soon as OMB clearance is received).

**District recruiter training and tracking.** To help establish a strong rapport between the study team and the district administrators and their support staff, each district will be assigned one primary recruiter. Recruiters will participate in a half-day training session covering the following topics: (1) communicating the study's importance, data collection needs, and methodological issues in effective, nontechnical terms, (2) describing the study, (3) procedures for district recruitment, (4) addressing potential recruitment issues, and (5) task management (including the recruiting schedule, weekly meetings, and use of the tracking system). Trainers will also share key strategies based on the contractor's extensive experience recruiting districts for other IES studies, including the Impact Evaluation of Race to the Top and School Improvement Grants.

**District research applications and MOUs.** We will gain the necessary approvals and abide by the relevant guidelines for conducting the study in the 14 selected districts. In districts with policies concerning external research projects, we will seek expedited research application reviews where possible and emphasize that the study has received prior review by its institutional review board (IRB).

Following district officials' oral commitment to participate in the study, they will be sent a memorandum of understanding (MOU) to sign. The MOU will describe the agreed-upon roles and responsibilities of the study team and of participating districts (Appendix C). A representative of the study contractor will also sign the MOU.

All contacts with districts will be documented in an electronic tracking system to minimize redundant contacts and to provide the Contracting Officer's Representative (COR) with up-to-date information on recruiting progress. All recruiters will have access to the web-based system.

## 2. Procedures for collection of information

### a. Statistical methodology for stratification and sample selection

We will not draw a random sample of states (or districts) for the study. The descriptive and feasibility components will include all nine states that received

RTT-ELC Round 1 grants. For the analysis of TQRIS ratings and child outcomes, states (and, if necessary, districts) will be selected purposefully, based on the availability of TQRIS and kindergarten entry assessment data and to maximize statistical power. The sample will consist of at least 200 center-based ECE programs from an estimated 2 to 3 of the 9 states that received RTT-ELC Round 1 grants. If kindergarten entry assessment data are not available from states, we will collect data from approximately 14 school districts within the 2 to 3 states, which will be selected purposefully based on the availability of kindergarten assessment data and to maximize statistical power.

### b. Estimation procedures

Our estimation procedures comprise two sets of analyses aligned with research question 3. These analyses will (1) describe the distribution of the TQRIS ratings and (2) estimate the association between the ratings and child outcomes.

**Descriptive statistics.** For the center-based ECE programs included in our analysis within a given RTT-ELC state, we will describe the distribution of the TQRIS ratings by providing a figure that includes the proportion of programs at each rating level. If the data are available, we will also provide descriptive statistics for the component-level ratings. We will describe the distribution of the component-level ratings by providing a table that includes the number of programs that received each rating level.

For the study programs included in the analysis in each RTT-ELC state, we will also describe the kindergarten entry assessment outcomes and the demographic characteristics of the children who attended the programs. We expect kindergarten entry assessments to measure several domains of school readiness (including cognition and general knowledge, language and literacy, and others). Hence, we will provide a table that includes the mean readiness scores (if an overall score is generated), the mean scores for each assessment domain, and the demographic characteristics of the children who attended programs at each rating level. We will describe the distributions of the kindergarten entry assessment scores by providing box plots or histograms.

These descriptive statistics will provide useful information about the construction of each state's TQRIS. If the rating system fails to differentiate among center-based ECE programs in the sense that most programs receive the same rating, the ratings will be unable to predict child outcomes. If none of the center-based ECE programs achieve the top rating for a particular TQRIS component, it may indicate that the standards are unrealistic or unattainable. As with the overall ratings, if there is little variation between programs in the

component-level ratings, then the ratings will not be able to adequately distinguish between programs of different levels of quality. States' collection of information on these TQRIS components (and their indicators) could still be useful for other state purposes, however. For example, states might still want to collect information about safety to ensure compliance with state regulations, even if there is little variation across programs in these measures. For quality indicators that very few programs achieve, states could use the data to provide training and technical assistance around those indicators.

**Association between child outcomes and overall TQRIS ratings.** To measure the association between kindergarten entry assessment scores and the overall TQRIS ratings, we will estimate the following regression model for each state in our sample:

$$(1)\quad Y_{ipt} = \alpha + \beta X_{p,t-1} + \varepsilon_{ipt}$$

where $Y_{ipt}$ is the overall or domain-level score[9] in year *t* for child *i* who attended program *p* in year *t-1*, $X_{p,t-1}$ is the overall TQRIS rating for program *p* in year *t-1*, $Z_{it}$ is a vector of controls for the child characteristics that are available in the administrative kindergarten data (for example, gender, race/ethnicity, free or reduced-price lunch status, English Language Learner status, special education status, and age),[10] and $\varepsilon_{ipt}$ is a random error term. We will standardize each of the kindergarten entry assessment scores $Y_{ipt}$ using the state- and year-specific means and standard deviations.[11] The coefficient $\beta$ measures the average change in student outcomes associated with a one-unit improvement in the overall TQRIS rating. Standard errors will be clustered by program to account for possible correlations between the outcomes of students who attended the same program. To account for the fact that we may be examining multiple child outcomes, we

[9] The kindergarten entry assessments will likely measure several of five school readiness domains: (1) language literacy, (2) cognition and general knowledge (early mathematics and early scientific development), (3) approaches toward learning, (4) physical well-being and motor development (including adaptive skills), and (5) social and emotional development.

[10] We will also estimate specifications that assess the sensitivity of the estimates to the inclusion of district fixed effects in $Z_{it}$. District fixed effects could help account for unobserved differences among the characteristics of children who attend different school districts.

[11] Standardizing the kindergarten entry assessment scores will achieve several goals. First, it will enable us to present findings in a scale that is more readily interpretable than the original assessment scale. Second, it will allow us to conduct a pooled analysis if we are able to obtain data for multiple cohorts of children within a state. Third, even though we will do separate analyses by state, standardization will help facilitate cross-state comparisons of the relationship between TQRIS and kindergarten entry assessment scores (because states may use different assessments).

will adjust *p*-values in these regressions using the Benjamini and Hochberg (1995) procedure, as did Zellman et al. (2008). Because we anticipate having only one cohort of kindergarteners in each RTT-ELC state, this equation does not include year indicators. We discuss the possibility of having multiple cohorts next.

***(1) Accounting for multiple cohorts***

The inclusion of multiple cohorts in the analyses could improve the precision of the estimated relationships. If kindergarten entry assessment and TQRIS data from earlier years are available, *and* if the kindergarten entry assessments and TQRIS standards used in those years were the *same* as those used in the focal year for the study (that is, fall 2015 kindergarten entry assessments and 2014–2015 TQRIS ratings), we will modify equation (1) as follows:

$$(2)\quad Y_{ipt} = \alpha + \beta X_{p,t-1} + \gamma Z_{it} + \lambda D_t + \varepsilon_{ipt}$$

where $D_t$ contains indicator variables for each year of the kindergarten entry assessment. The inclusion of these year fixed effects accounts for the average differences in test scores across years that result from changes not related to TQRIS (for example, changes in the difficulty of the tests). Because our sample will be restricted to kindergarteners who attended a center-based ECE program in year *t-1,* the year fixed effects also control for differences in the average ratings of programs across years that are not the result of changes in program quality (for example, differences in the stringency of the raters across years on a an observational rating component).[12]

As a sensitivity analysis, we will include interactions between the ratings and the year fixed effects in Equation (2) to test for changes over time in the relationship between ratings and child outcomes. It is possible that this relationship could change over time due to factors other than program quality, such as differential selection of the programs participating in the rating system. If we find evidence that the relationships between the ratings and child outcomes change over time, we will conduct separate analyses for each cohort of children.

***(2) Addressing non-linear relationships***

Because the relationship between child outcomes and changes in the TQRIS rating could be nonlinear across the ratings distribution—for example, the effect

[12] If states have established inter-rater reliability of the ratings, differences in the stringency of raters may be less of a concern. Nevertheless, the year fixed effects would account for other factors unrelated to program quality that may affect ratings from year to year.

of moving from Level 1 to Level 2 may be different than the effect of moving from Level 3 to Level 4—we will also estimate variants of Equation (1) that replace programs' TQRIS rating $X_{p,t-1}$ with a vector of indicator variables for each of the quality ratings. We will report the F-statistic from a joint hypothesis test that the coefficients on the indicators are linear combinations of each other, which Equation (1) assumes. If we find evidence of nonlinear relationships—that is, if the F-statistic for the joint hypothesis test is statistically significant—we will include these indicator variables in all future regression analyses.

**Association between child outcomes and component-level TQRIS ratings.** If data on the component- and indicator-level TQRIS ratings are available, we will measure the association between kindergarten entry assessment scores and these ratings. To achieve this goal, we will estimate the regression model in Equation (1), using the component- and indicator-level TQRIS rating for program *p* in year *t-1* as $X_{p,t-1}$.[13] This analysis may help uncover which components of the TQRIS are more strongly related to child outcomes and which are less strongly related. This information could help states refine and improve their TQRIS systems.

**Study limitations.** Our analysis of the relationship between child outcomes and the TQRIS ratings has several limitations (described further below), including that (1) estimates cannot be interpreted as causal effects of attending a high-quality center-based ECE program, (2) it is limited to the domains measured by the kindergarten entry assessments at a single point in time, (3) it is limited to the measures of program quality that are included in the TQRIS of the sample states, and (4) it may not necessarily generalize to other contexts.

The first limitation is that because students are not randomly assigned to programs, these estimates cannot be interpreted as causal effects of attending a high-quality center-based ECE program. A positive relationship between the TQRIS ratings and kindergarten entry assessment scores may be due to highly rated programs having more positive effects on children or due to differences in the characteristics or circumstances of children whose parents *choose* highly rated programs. For example, a program located in a high-income neighborhood might seem to produce higher levels of school readiness if it enrolls mostly advantaged children from the neighborhood. To the extent that data are available, we will examine and describe differences in the characteristics of children attending programs that receive different ratings.

[13] If it is not possible to include all component-level ratings in a single model due to multicollinearity, we will fit separate models for each component.

Although we can control for the characteristics of children that are observable in the data, there may be other unobserved factors affecting child outcomes that could be correlated with TQRIS program ratings, for example, parental involvement. Therefore, when interpreting findings, we will clarify that the outcome differences (even after controlling for observable child characteristics) could still reflect unobserved differences between children attending the programs and not necessarily differences in program quality.

Even if our estimates were unbiased, it would still be important to not overstate the value of the TQRIS ratings, because we do not know whether the ratings in fact account for all important aspects of quality. Parents and educators may have access to types of information about the quality of programs other than what is captured through the TQRIS ratings (for example, from referrals from other parents or through their direct interactions with program staff).

A second limitation is that child outcomes will be limited to the domains measured by the kindergarten entry assessments at a single point in time, and may not represent all of the outcomes that are important for child development. The relationship between the quality ratings and other unobserved but important child outcomes (including growth over time) might differ from the relationship between the ratings and the outcomes we observe.

A third limitation is that we observe only the measures of program quality that are included in the TQRIS of each state in our sample. Other measures of quality—which might be included in other state systems—could be more highly correlated with child outcomes.

A fourth limitation concerns external validity. Our results will not necessarily generalize to other contexts because we will examine quality ratings for center-based ECE programs in only two or three Round 1 RTT-ELC states. Because these states received RTT-ELC grants, their TQRIS may be better established than those in other states. It is also possible that the programs in our sample may not be representative of all center-based ECE programs, or even of all center-based ECE programs that participate in TQRIS. Our results also will not generalize to other types of child care programs participating in TQRIS but not included in the analysis.

### c. Degree of accuracy

We propose that the minimum detectable correlation (MDC) for this study would ideally be low enough to yield a difference in child outcomes between the highest and lowest quality programs in the sample of 0.10 standard deviations

(though the sample would include programs at all quality levels), and that the MDC should be no larger than the level necessary to yield a difference in child outcomes of 0.20 standard deviations. We looked at findings from the final impact study of Head Start (ACF 2010) for guidance in selecting a target MDC. In that study, significant impacts of Head Start on kindergarten and first grade outcomes were generally in the range of 0.10 and 0.20 standard deviations. We infer from these findings that it is unlikely that the outcomes for children in the highest quality programs would be any more than 0.20 standard deviations greater than the outcomes for children in the lowest quality programs, and that a 0.10 standard deviations difference may be more plausible.

We anticipate that both of our proposed data collection approaches—collecting data from states or individual school districts—would yield sufficient statistical power to detect meaningful relationships between child outcomes and quality ratings. We refer to the first of these approaches as the untargeted sample, because data would be collected on all center-based ECE programs within a state. We refer to the second of these approaches as the targeted sample, because data would be collected from a targeted sample of school districts within a state. We may be able to collect data from an untargeted sample of programs in one state but still need to collect data on a targeted sample in another state.

**Preferred approach: Collecting data from states.** Obtaining kindergarten entry assessment data from states is our preferred approach, as it would afford greater statistical power to detect the relationship between TQRIS ratings and child outcomes (if one exists). It would also enable us to look (with an acceptable level of precision) not just at the difference between the highest and lowest rating categories but also compare programs at the intermediate levels. We anticipate that, for each state, it will be possible to detect an MDC of 0.03 between child outcomes and quality ratings using an untargeted sample of at least 1,000 center-based ECE programs, an MDC of 0.04 using an untargeted sample of 500 center-based ECE programs, and an MDC of 0.06 using an untargeted sample of 200 center-based ECE programs. These MDCs correspond to differences in child outcomes between the highest and lowest quality center-based ECE programs of 0.10, 0.14, and 0.21 standard deviations, respectively.

In Table B.1, we present MDC values for different sample sizes, and the difference in child outcomes between the highest and lowest quality programs (represented by the top and bottom 10 percent of the ratings distribution).

**Table B.1. Minimum Detectable Correlations, Untargeted Sample by State**

| | Minimum Detectable Correlation Between Quality Rating and Child | Difference in Child Outcome Between Top and Bottom 10 |
|---|---|---|

| | Outcome | Percent of Program Quality Distribution |
|---|---|---|
| Intraclass correlation = 0.05 | | |
| 100 programs | 0.08 | 0.27 |
| 200 programs | 0.06 | 0.21 |
| 500 programs | 0.04 | 0.14 |
| 1,000 programs | 0.03 | 0.10 |
| Intraclass correlation = 0.10 | | |
| 100 programs | 0.10 | 0.34 |
| 200 programs | 0.07 | 0.25 |
| 500 programs | 0.05 | 0.14 |
| 1,000 programs | 0.03 | 0.11 |

Note: The second column shows the smallest correlation between a child outcome and a quality rating that can be detected with 80 percent probability, assuming two-tailed hypothesis testing, a 5 percent significance level, 30 pre-K students per program, and the intraclass correlation (ICC) and sample sizes indicated in the first column of the table. The correlation is calculated at the child level, but the quality rating varies only at the program level. The third column shows the difference in child outcomes between the top and bottom 10 percent of programs (ranked by quality rating) that corresponds to the correlation reported in column 2. The difference reported in the third column is standardized using the student-level standard deviation of the student outcome. All values in the table were calculated using a simulation where data were generated assuming the ICC, sample size, power, and significance levels stated above. We also assumed normality of the child outcome and the quality rating. The simulation was run with 10,000 replications. These values do not include an adjustment for multiple comparisons.

**Alternative approach: Collecting data from districts.** If TQRIS data are available from states but child outcomes data are not, we will try to collect child outcomes data from individual school districts. Under this scenario, for a given state, we might still be able to detect a correlation of 0.06 between child outcomes and quality ratings and a difference of around 0.18 standard deviations between the highest and lowest quality programs, using a sample of roughly 200 programs across several carefully selected districts. We would select districts with the greatest variability in TQRIS ratings (and a more uniform distribution across the rating levels) to maximize the number of programs in our sample at the top and bottom of the ratings distribution. However, we would still collect data for all programs in the selected districts and also analyze the relationship between child outcomes and TQRIS ratings for programs at all ratings levels. Under this approach, the statistical power for that analysis will be similar compared with a scenario whereby we are able to obtain data from states for a much larger number of programs, but our selective approach to choosing districts can help maximize power to detect differences between the highest and lowest quality programs.

In Table B.2, we report minimum detectable differences (MDDs) between the highest and lowest quality programs, assuming that we are successful in targeting districts with a larger number of programs at the top and bottom of the ratings distribution. So, for example, assuming an intraclass correlation of 0.05, the MDD

would be 0.18 given a total sample of 200 programs if 40 of those 200 programs were in the top level and 40 were in the bottom level in a state with a 5-level TQRIS: the corresponding MDC if we included all 200 programs in the sample would be 0.06. The estimated MDDs shown in Table IV.2 are lower than those shown in Table IV.1 because the targeted sample would contain more programs from the top and bottom rating categories as compared to the untargeted sample.

**Table B.2. Minimum Detectable Differences, Targeted Sample by State**

| | Minimum Detectable Correlation Between Quality Ratings and Child Outcomes in the Full Targeted Sample | Difference in Child Outcomes Between Programs in the Top 20 Percent and Bottom 20 Percent of the Quality Rating Distribution |
|---|---|---|
| Intraclass correlation = 0.05 | | |
| 50 programs | 0.11 | 0.38 |
| 100 programs | 0.08 | 0.26 |
| 200 programs | 0.06 | 0.18 |
| 500 programs | 0.04 | 0.11 |
| Intraclass correlation = 0.10 | | |
| 50 programs | 0.15 | 0.48 |
| 100 programs | 0.10 | 0.33 |
| 200 programs | 0.07 | 0.23 |
| 500 programs | 0.05 | 0.14 |

Note: This table shows the smallest correlation and difference between child outcomes and program quality ratings that can be detected with 80 percent probability, assuming two-tailed hypothesis testing, a 5 percent significance level, 30 pre-K students per program, and the intraclass correlation (ICC) and sample sizes indicated in the first column of the table, assuming a uniform distribution of quality ratings among programs and a linear relationship between child outcomes and quality ratings. These values do not include an adjustment for multiple comparisons. The values reported in the second column correspond to an analysis that calculates the difference in means between two groups (programs with the highest and lowest quality ratings); the values reported in the first column correspond to calculating a correlation between two continuous variables.

## 3. Methods to maximize response rates and deal with nonresponse

A clear description of the study design presented to school districts, and the data collection's reliance on administrative data without any direct data collection from students or intrusion on classroom instructional time, will encourage cooperation with study efforts. Based on recruiting rates from prior IES studies, we anticipate a response rate of about one-third, or 33%. Thus, we will need to reach out to about 42 districts to ultimately recruit 14 into the study.

Following initial mailings of the district notification letter (Appendix A) and the non-technical study information sheet (Appendix B) to district superintendents, recruiters will make follow up calls to the district contacts to introduce themselves and the study, and to arrange a time to further discuss the study. We will enlist the help of state representatives in our district recruitment

efforts by asking them to reach out to districts to introduce the study and follow-up on our behalf with non-responsive districts, if necessary. Recruiters will make multiple attempts within a 2-3 week period to reach district contacts by telephone or email to initiate a dialog about the study. In past recruiting efforts, Mathematica recruiters have been successful in establishing contact with the majority, if not all, district representatives to initiate discussions about the study design and planned data collection activities. As such, we expect all 42 districts to participate in recruiting discussions (a response rate of 100 percent). After the initial discussion, recruiters will remain in contact with the district representative, again by telephone and email as needed, to address questions. A sample template for email follow-up is included in Appendix D. Recruiters will continue to reach out to district representatives until an agreement to participate is reached or a definitive negative response is given. We anticipate that 14 out of 42 districts will agree to participate following recruitment discussions. We will request data from these 14 districts using the data request form in Appendix E. We expect all 14 districts to complete the data request (a response rate of 100 percent).

Reducing districts' burden in the submission of study data will facilitate a high response rate for the administrative data collection. Federal rules permit ED and its designated agents to collect student demographic and existing achievement data from schools and districts without prior parental or student consent (Family Educational and Rights and Privacy Act [20 U.S.C. 1232g; 34 CFR Part 99]). To maximize the response rate and minimize burden on school districts, we will follow these federal rules.

## 4. Pilot testing

We will not conduct any pilot testing for the recruitment or data collection phases of the study.

## 5. Individuals consulted on statistical aspects of the design

The following individuals were consulted on the statistical aspects of this study:

| Name | Title | Telephone |
|---|---|---|
| Irma Perez-Johnson | Associate Director of Research, Mathematica (Study Director) | 609-275-2339 |
| Susanne James-Burdumy | Senior Fellow and Education Area Leader, Mathematica | 609-275-2248 |
| John Deke | Senior Researcher, Mathematica | 609-275-2230 |

| Name | Title | Telephone |
|---|---|---|
| Lisa Dragoset | Senior Researcher, Mathematica | 609-945-3348 |
| Hanley Chiang | Researcher, Mathematica | 617- 674-8374 |

The following individuals will be responsible for data collection and analysis:

| Name | Title | Telephone |
|---|---|---|
| Irma Perez-Johnson | Associate Director of Research, Mathematica (Study Director) | 609-275-2339 |
| Susanne James-Burdumy | Senior Fellow and Education Area Leader, Mathematica | 609-275-2248 |
| Gretchen Kirby | Senior Researcher, Mathematica | 202-484-3470 |
| Pia Caronongan | Researcher, Mathematica | 617- 674-8368 |
| Mariesa Herrmann | Researcher, Mathematica | 609-716-4544 |

## REFERENCES

Administration for Children and Families, U.S. Department of Health and Human Services. "Head Start Impact Study." Final Report. Washington, DC: ACF, January 2010.

Benjamini, Y., and Y. Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series* B *(Methodological),* vol. 57, no. 1, 1995, pp. 289–300.

Elicker, J., C.C. Langill, K. Ruprecht, J. Lewsader, and T. Anderson. "Evaluation of 'Paths to QUALITY,' Indiana's Child Care Quality Rating and Improvement System: Final Report (Technical Report #3)." Department of Human Development and Family Studies, Center for Families, Purdue University, 2011.

Lahti, M., T. Sabol, R. Starr, C. Langill, and K. Tout. "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, 2013.

Lugo-Gil, J., S. Sattar, C. Ross, K. Boller, K. Tout, and G. Kirby.. "The Quality Rating and Improvement System (QRIS) Evaluation Toolkit." OPRE Report# 2011-31. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2011.

Thornburg, K.R., W.A. Mayfield, J.S. Hawks, and K.L. Fuger. "The Missouri Quality Rating System School Readiness Study." Columbia, MO: University of Missouri Center for Family Policy and Research, 2009. Retrieved from http://mucentermissouri.edu/MOQRSreport.pdf.

Tout, K., R. Starr, T. Isner, J. Cleveland, L. Albertson-Junkans, M. Soli, and K. Quinn. "Evaluation of Parent Aware: Minnesota's Quality Rating and Improvement System Pilot." Washington, DC: Child Trends, 2011.

Zellman, G.L., and R. Fiene. "Validation of Quality Rating and Improvement Systems for Early Care and Education and School-Age Care." Research-to-Policy, Research-to–Practice Brief OPRE 2012-29. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2012.

Zellman, G.L., M. Perlman, V. Le, and C. Setodji. "Assessing the Validity of the Qualistar Early Learning Quality Rating and Improvement System as a Tool for Improving Child-Care Quality." Pittsburgh, PA: Rand Corporation, 2008. Accessed November 26, 2012 at http://www.rand.org/pubcs/monographs/MG650.html.

MATHEMATICA
Policy Research

**www.mathematica-mpr.com**

Improving public well-being by conducting high quality, objective research and surveys

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC