This product is part of the RAND Corporation documented briefing series. RAND documented briefings are based on research briefed to a client, sponsor, or targeted audience and provide additional information on a specific topic. Although documented briefings have been peer reviewed, they are not expected to be comprehensive and may present preliminary findings.

DOCUMENTED
BRIEFING

How Would Programs Rate Under California's Proposed Quality Rating and Improvement System?

Evidence from Statewide and County Data on Early Care and Education Program Quality

Lynn A. Karoly, Gail L. Zellman

Sponsored by the David and Lucile Packard Foundation

RAND EDUCATION and
RAND LABOR AND POPULATION

This study was sponsored by the David and Lucile Packard Foundation and was conducted jointly in RAND Education and RAND Labor and Population, units of the RAND Corporation.

Published 2012 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

In 2008, California Senate Bill (SB) 1629 established the California Early Learning Quality Improvement System (CAEL QIS) Advisory Committee, which was tasked with developing the policy and implementation plan for an early care and education (ECE) quality rating and improvement system (QRIS). The aim of the CAEL QIS Advisory Committee in designing a QRIS was to improve the quality of ECE programs serving children from birth to kindergarten entry. The advisory committee's final report, which included a recommended QRIS design, was published in December 2010.

In November 2009, Governor Arnold Schwarzenegger issued Executive Order S-23-09 establishing the California State Advisory Council on Early Learning (also known as the ELAC, Early Learning Advisory Council). The ELAC was to build upon the quality rating system work of the CAEL QIS Advisory Committee after its design effort was completed, and the ELAC was tasked with developing future policy for early learning in the state.

The CAEL QIS Advisory Committee understood that the successful development of an effective and robust QRIS for California would require the testing of various system elements, as well as the system as a whole, through a series of pilot studies. However, it will take time for field-based pilot work to get under way and produce results. In this early work, we use existing statewide and county data on ECE program quality to assess the distribution of ECE programs across the rating tiers for different quality elements in the proposed QRIS, as well as combinations of quality elements.

The goal of the modeling effort was to provide the ELAC members and other stakeholders with information that could inform the pilot phase in the development of a California QRIS. The work also suggests a method that QRIS designers and implementers in other states might use to guide their design process, to assess whether ratings are working as planned, and to focus their quality improvement efforts. For this reason, in addition to stakeholders in the development of a QRIS in California, this study may be of interest to policymakers, agency staff, ECE providers, and child advocates involved in the design and implementation of QRISs in other states.

This study was funded by the David and Lucile Packard Foundation as part of its support for the RAND Corporation's assistance to the state of California's efforts to develop, pilot, implement, and evaluate a QRIS. This research was conducted jointly in RAND Education and RAND Labor and Population, units of the RAND Corporation. For inquiries related to RAND Education, please contact Darleen Opfer, Director, RAND Education, at Darleen_Opfer@rand.org. For inquiries related to RAND Labor and Population, please contact Arie Kapteyn, Director, RAND Labor and Population, at Arie_Kapteyn@rand.org.

Table of Contents

Summary

In 2010, the California Early Learning Quality Improvement System (CAEL QIS) Advisory Committee recommended a structure for a voluntary quality rating and improvement system (QRIS) that could apply to the state's 11,000 licensed centers and 36,600 licensed family child care homes (FCCHs). The proposed design consisted of an unweighted block system with five tiers, in which all quality criteria in each tier must be met in order to obtain a rating at that tier. The rating structure provided for five quality elements: ratio and group size, teaching and learning, family involvement, staff education and training, and program leadership.

The aim of this study was to conduct an initial examination of some key aspects of the proposed QRIS design. By capitalizing on two existing data sets that included several of the quality elements in the proposed QRIS rating design, the work highlights some relationships among these quality elements, examines the ways in which different measures of these elements proposed in the QRIS design relate to each other, and provides information about the likely distribution of program-level ratings across the proposed rating tiers. This work provides California QRIS planners and other stakeholders with important information about some fundamentals of the proposed QRIS rating scheme that could inform California's QRIS design in advance of field-based pilot efforts.

Study Approach

The study relied on two existing data sources to achieve two primary objectives:

- Assess the likely distribution of early care and education (ECE) programs across the five rating tiers in the QRIS proposed by the CAEL QIS Advisory Committee.
- Identify the implications of the findings for California's QRIS design and for future pilot studies.

To fully analyze the proposed QRIS, the ideal data source would include information on all five quality elements for a representative sample of center-based providers and FCCHs serving infants, toddlers, and preschool-age children. The two sources of data available to us provide much, but not all, of this information:

- The first includes a representative statewide sample of center-based programs serving preschool-age children observed as part of the RAND California Preschool Study, with measures of program quality corresponding to three of the five quality elements in California's proposed rating structure.
- The second includes a set of programs serving infants, toddlers, and preschool-age children in centers and FCCHs that participated in San Francisco County's Gateway to Quality (GTQ) initiative, with observation-based data on one of the five proposed QRIS rating elements.

The RAND statewide data consist of a set of providers that is more representative of all providers in the state than the set of providers that might be expected to

participate in a voluntary QRIS would be; voluntary QRISs tend to attract higher-quality providers. The data do not include FCCH providers or programs serving infants and toddlers. The GTQ data may better reflect the set of center- and home-based providers that would agree to participate in a QRIS targeted at programs receiving subsidies, as well as the effect of technical assistance (TA) on quality. Neither data source allows us to fully estimate the distribution of ratings under the proposed QRIS because we do not have the quality measures associated with all five quality elements. Despite these limitations, the two data sources provide information about the range of provider types expected to be covered by the proposed QRIS and allow us to examine key dimensions of quality that are likely to be included in a rating system for California.

Following the study objectives, these two sources of data were used first to determine the distribution of single quality elements in the QRIS and then the distribution of multiple quality elements taken together, the latter to replicate the QRIS as closely as possible. The data were also used to determine the expected share of programs that would fall into each tier of the rating system given specified cut points. The analyses further examined the relationship among different quality elements and pinpointed the elements that are likely to suppress a program's ability to improve its rating in the block system that underlies the proposed CAEL QIS design.

Key Findings Regarding Provider Ratings

Our analysis of the statewide data set, which pertains to a representative sample of center-based programs serving preschool-age children, demonstrates the following:

- **Given the proposed cut points in the QRIS design, center-based programs would score better on some quality elements than others.** Most programs (upward of 80 percent) would reach Tiers 3 to 5 (out of the proposed five tiers) on the ratio and group size quality element. About half of programs would reach Tier 4 or 5 for the staff education and training quality element where the proposed standard is an associate's degree (Tier 4) or bachelor's degree (Tier 5). In contrast, just one in four programs would reach Tier 4 or 5 based on the teaching and learning quality element, measured by either the Early Childhood Environment Rating Scale–Revised (ECERS-R) or the Classroom Assessment Scoring System (CLASS).
- **The teaching and learning element would constrain programs from reaching higher rating tiers more than the ratio and group size or the staff education and training elements.** Expectations in the design phase that quality elements will be highly correlated may not always hold. We found, for example, that having staff with levels of education that qualify for the highest tiers does not necessarily ensure that a program will score at that same level on the relevant environment rating scale (ERS) or on the CLASS.
- **Overall, given the proposed cut points and the three quality elements examined, few programs would initially reach the highest rating tiers.** Our

estimate that fewer than 10 percent of center-based programs would reach the top tier (Tier 5) is consistent with the CAEL QIS Advisory Committee's goal of making the upper tiers aspirational.

Findings from the GTQ initiative data set, which includes programs that received various types of state or local public funding, illuminate the distribution of program-level ratings that might result when certain programs are targeted, as well as the likely variation in ratings across provider types and by the ages of children served. In particular, we found the following:

- **Compared with the universe of providers, program-level ratings may be higher if the QRIS targets programs with public funding.** More center-based programs in GTQ—in which programs receiving public subsidies are targeted—would score high enough on the ECERS-R to reach Tier 4 compared with the programs included in the statewide data—which contain a representative sample of all providers. This difference may reflect the selectivity of the programs in GTQ, which includes publicly funded programs with higher quality standards, such as California State Preschool and other Title 5 programs, as well as Head Start. Quality may also be higher in the GTQ sample as a result of the quality improvement supports that GTQ provides. Even so, like the statewide data, few GTQ programs would reach Tier 5.
- **Quality ratings are likely to vary by the ages of children served and the setting type.** In GTQ, center-based classrooms serving infants and toddlers scored somewhat lower on the Infant/Toddler Environment Rating Scale–Revised (ITERS-R) than rooms serving preschool-age children evaluated using the ECERS-R. FCCHs evaluated using the Family Child Care Environment Rating Scale–Revised (FCCERS-R) scored considerably lower than center-based programs scored on the ECERS-R.

It is important to note that the available data did not allow us to fully replicate the ratings that would result if the QRIS proposed by the CAEL QIS Advisory Committee were implemented. On the one hand, if all providers were rated (i.e., no selectivity), our estimates of rating distributions across tiers based on the statewide sample of center-based programs can be viewed as an upper bound on how programs would likely be distributed. This is because, in a block system, programs cannot be rated any higher on all five quality elements than they score on the three elements we considered. On the other hand, because it is reasonable to assume that only those programs with higher quality or higher ambitions would choose to participate in a voluntary system, the share of programs in the upper rating tiers in such a system likely would be higher than our statewide estimates. The share in the upper tiers would also be expected to increase over time as supports for quality improvement embedded in the proposed QRIS design help programs improve their quality.

Implications for California's Quality Rating and Improvement System Design and Future Pilot Efforts

These findings regarding program ratings under the proposed QRIS have implications for the design of the QRIS, as well as for future pilot efforts. With respect to design, the findings show that the proposed design is likely to result in a small share of programs initially reaching the highest rating tiers; this allows more room for quality improvement over time, although it may risk lower participation if programs believe they are unlikely to achieve the highest rating. The results point to the value of standards. Programs currently score highest on those quality elements for which standards are already in place (e.g., ratio and group size) and suggest that the broader set of standards in the proposed QRIS would lead programs to improve on the other QRIS elements as well. The data also suggest that programs are likely to need the most assistance with improving their scores on the relevant ERSs (e.g., ITERS-R, ECERS-R, or FCCERS-R and the CLASS if it is included). Another concern is how to generate meaningful center-level ratings for programs for which rating elements may vary across classrooms serving children in the same age group or across classrooms serving younger versus older children in the birth-to-five age range.

The study findings point to the value of using existing survey or administrative data sets in advance of more-extensive and more-costly field pilots to examine key aspects of a QRIS rating scheme. The statewide and county data examined here reveal the degree to which different quality elements are related (e.g., which rating elements may constrain program ratings more than others) and how closely the assumptions made by designers accord with the realities of how programs operate (e.g., that high levels of staff education and training will necessarily be associated with high ratings on a given ERS). However, the data available for this study did not measure several key quality elements included in the California QRIS design—namely, measures of family involvement and program leadership. It would be important to launch an early pilot effort to measure those quality elements and assess the implications of their inclusion in the rating system. Piloting the proposed QRIS as a whole would also provide more information about the likely distribution of programs in a voluntary system or one that targets certain providers (e.g., those receiving public subsidies).

Acknowledgments

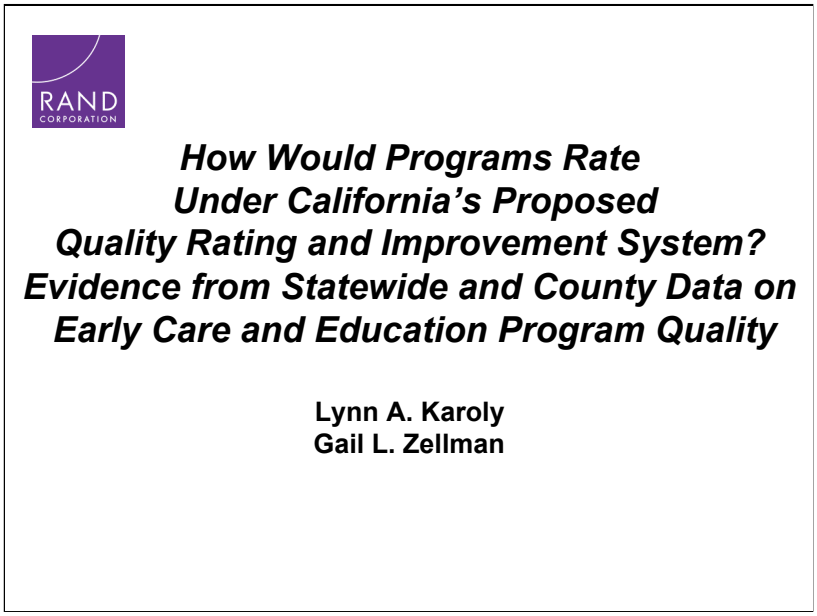
We are indebted to Meera Mani at the Packard Foundation who provided steady guidance and support throughout the course of this project. We are also grateful for the access to the San Francisco Gateway to Quality (GTQ) data provided by Gretchen Ames, Project Coordinator for GTQ, and her colleagues. In addition, we would like to acknowledge the very able programming support provided by our RAND colleagues Adria Jewell and Gerald Hunter.

The RAND Labor and Population review process employs anonymous peer reviewers, including at least one reviewer who is external to RAND. We benefited from the thorough and constructive feedback provided by two anonymous reviewers.

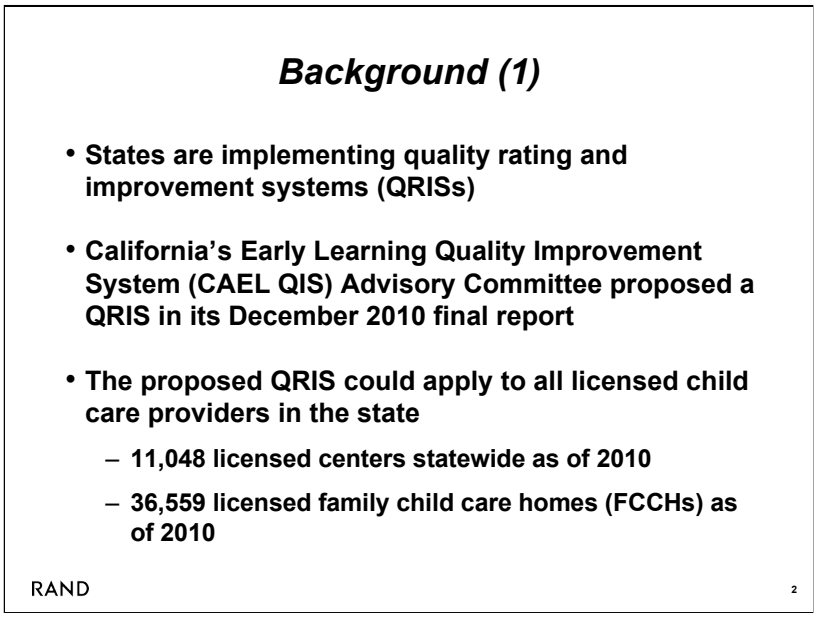
Abbreviations

AA	associate degree
BA	bachelor's degree
CAEL QIS	California Early Learning Quality Improvement System
CDA	Child Development Associate (credential)
CLASS	Classroom Assessment Scoring System
CO	classroom organization
ECE	early care and education
ECERS-R	Early Childhood Environment Rating Scale–Revised
ELAC	Early Learning Advisory Council
ERS	environment rating scale
ES	emotional support
FCCH	family child care home
FCCERS-R	Family Child Care Environment Rating Scale–Revised
GE	general education
GTQ	Gateway to Quality
IS	instructional support
ITERS-R	Infant/Toddler Environment Rating Scale-Revised
NAEYC	National Association for the Education of Young Children
NCCIC	National Child Care Information and Technical Assistance Center
NIEER	National Institute for Early Education Research
PD	professional development
PFA	Preschool for All
QI	quality improvement
QRIS	quality rating and improvement system
QRS	quality rating system
SB	Senate Bill
SE	student engagement

TA	technical assistance
UPK	universal prekindergarten program

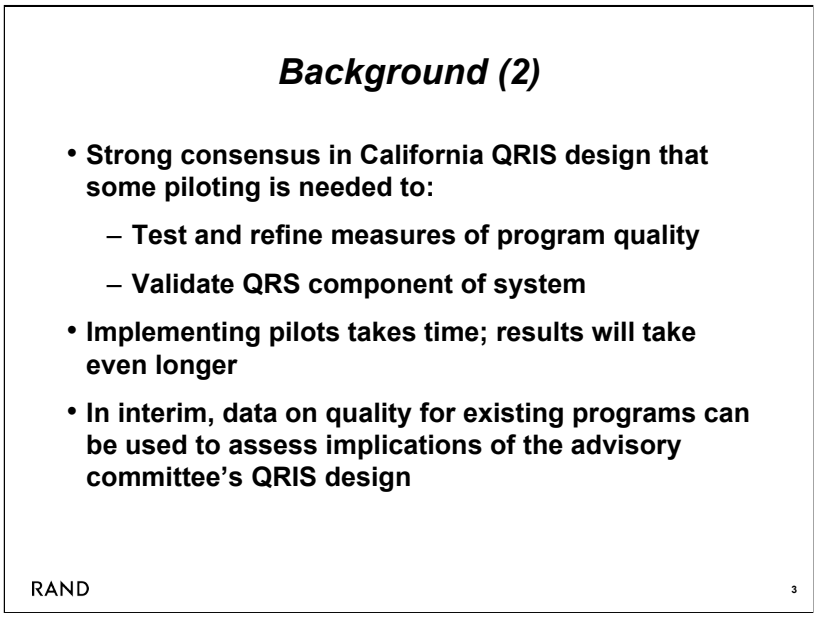


This presentation summarizes the results of an analysis of existing statewide and county data on early care and education (ECE) program quality to assess the distribution of ECE programs across the tiers of the quality rating and improvement system (QRIS) proposed by the California Early Learning Quality Improvement System (CAEL QIS) Advisory Committee in 2010.



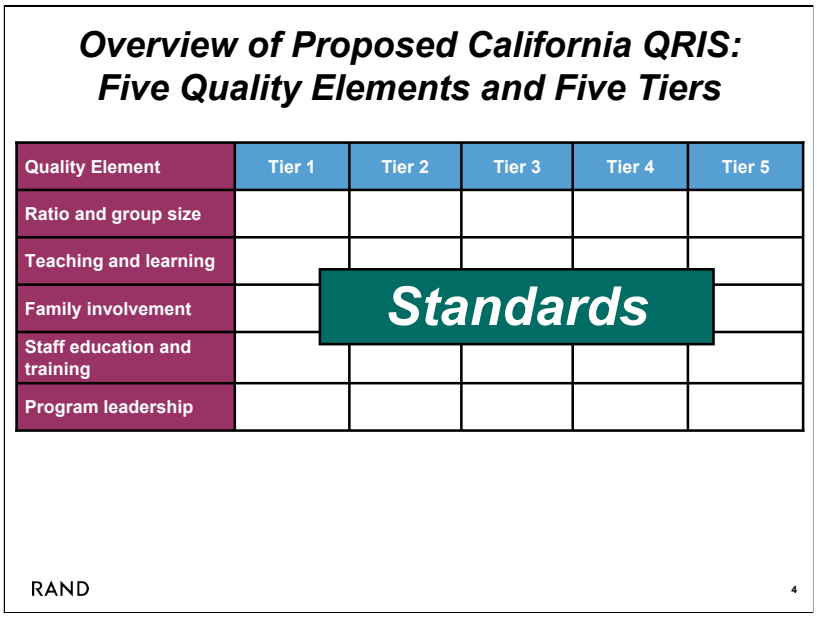
In an effort to raise the quality of ECE programs, more and more states have designed and implemented QRISs, multicomponent assessments of ECE programs designed to make quality transparent and easily understood (National Child Care Information and Technical Assistance Center [NCCIC], 2010). Participating providers are assessed on each of the system components and receive a summary rating that they are encouraged to display and that may be made public in other ways as well. In theory, these simple ratings (often zero to five stars or a rating of 1 to 4), enable parents, funders, and other stakeholders to make more-informed choices about which providers to use and support, and they encourage providers to improve the quality of care that their programs provide (Zellman and Perlman, 2008). QRISs also include feedback, technical assistance (TA), and other supports to motivate and support quality improvement (QI).

The proposed design for a California QRIS, developed by the CAEL QIS Advisory Committee and published in 2010, follows the path set by other states for such systems (CAEL QIS Advisory Committee, 2010a, 2010b). Like other states, the CAEL QIS Advisory Committee envisioned a voluntary QRIS that could potentially cover all types of licensed providers, although the advisory committee did not make an explicit recommendation regarding which providers would be included. According to data from the California Child Care Resource and Referral Network (2011), as of 2010, California had just over 11,000 licensed centers across the state and nearly 36,600 licensed family child care homes (FCCHs). Thus, a statewide QRIS could potentially reach nearly 47,000 providers statewide.



The successful development of an effective and robust QRIS for California will require the testing of its various elements, as well as the system as a whole, through a series of pilots. Indeed, as the work of the CAEL QIS Advisory Committee concluded and the design for the system took form, the importance of conducting pilot work on a range of system elements was repeatedly noted. Most other states that have developed a QRIS, or are currently in the process of developing one, have included a pilot phase. Pilots may be used to test and refine elements of the system, such as specific measures of program quality or the validity of the entire quality rating system (QRS) component of the QRIS. Pilots may also be used to test the feasibility of implementing the QRIS at scale, as well as to assess the associated costs of the system.

However, because the process of piloting a statewide QRIS can take months, if not years, to implement and even longer to generate results that will inform the final design and implementation, it is important to take advantage of any existing information that can shed light on the QRIS design. This study exploits statewide data previously collected by RAND on the quality of center-based programs in California serving preschool-age children (Karoly et al., 2008), as well similar data collected in San Francisco County for a set of programs participating in the Gateway to Quality (GTQ) initiative.

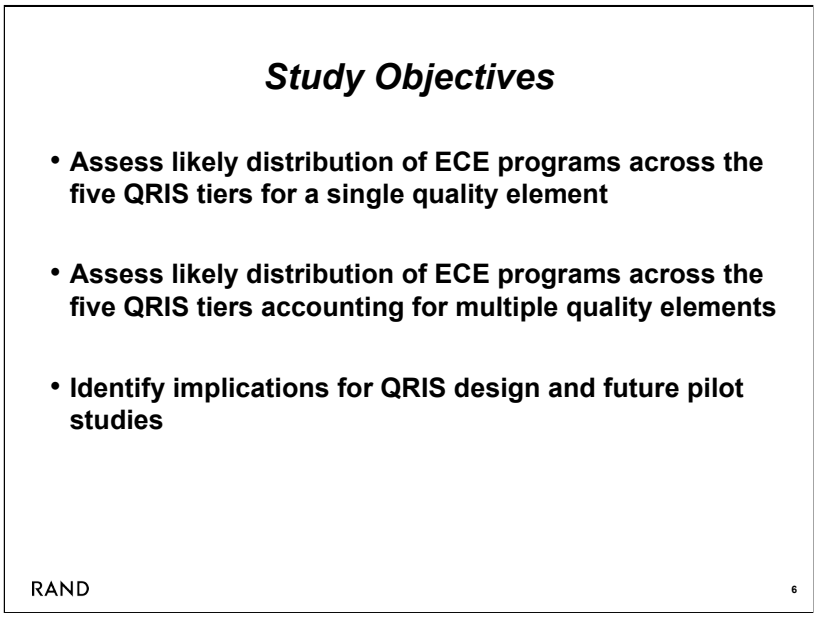


The proposed California QRIS establishes a set of quality standards for rating ECE programs, both centers and FCCHs, that would be included in the system. Specifically, the CAEL QIS Advisory Committee report recommended an unweighted block system for the rating structure's basic design (CAEL QIS Advisory Committee, 2010a, 2010b; see Tout et al., 2010, for a discussion of such systems). In a block system, all the quality criteria in each tier (i.e., in the matrix column for that tier) need to be met in order to obtain that rating. Thus, the rating is determined by the lowest tier attained across the quality elements. Moreover, as designed, each quality element receives the same weight in the summary rating.

The advisory committee approved five quality elements for the rating structure that are presented in the rows of the matrix in this chart: ratio and group size, teaching and learning, family involvement, staff education and training, and program leadership. These elements were selected because they were viewed as having the strongest evidence base, from either research or best practices, as factors that promote healthy child development in ECE settings. The advisory committee opted to focus on a relatively small number of critical quality elements in order to reduce the cost of implementing the rating process.

In addition, the advisory committee approved five tiers for each quality element, presented in the columns in the matrix, for which the criteria included in each block in each tier (i.e., each matrix cell) build on those in previous tiers for that block. Tier 1, with some modifications, tracks the state's Title 22 ECE program licensing standards. Tier 3 is consistent with the California Title 5 child development program standards, which apply to the State Preschool program and other Title 5 programs (see Karoly, Reardon, and Cho, 2007, for details). Finally, Tier 5 is modeled on nationally recommended ECE quality standards, including those embodied in the National Association for the Education of Young Children (NAEYC) accreditation standards and in the state preschool program benchmarks promulgated by the National Institute

for Early Education Research (NIEER) (NAEYC, 2007; Barnett et al., 2010). Notably, the top tier represents an aspirational level of quality; the advisory committee expected that only a small number of programs would attain the highest tiers when the QRIS begins. Over time, it is hoped, the QRIS will incentivize and support centers and FCCHs in moving to higher ratings (CAEL QIS Advisory Committee, 2010a, 2010b).

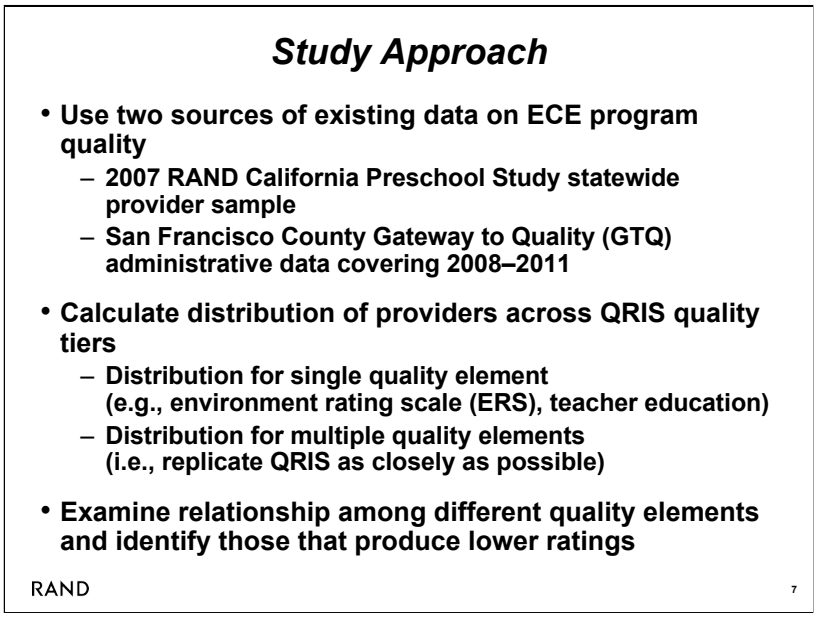


Broadly construed, the objective of this study was to use existing data on ECE program quality in California to assess the likely implications that the proposed QRIS design could have for how programs in the state would be rated.

The first goal was to examine how programs would be rated according to a given quality element and the proposed cut points for each rating tier. With the available data, discussed next, we could examine three of the elements in the proposed QRIS: ratio and group size, teaching and learning, and staff education and training. Examining a single element at a time shows whether programs currently rate relatively high or low on each element.

Our second aim was to calculate the distribution of programs across the rating scale when multiple elements are included. This allows us to more closely replicate the full rating system, in which standards for multiple dimensions of quality must be met at the same time. This analysis also serves to illustrate the extent to which different quality elements are correlated with one another and to identify quality elements on which otherwise well-rated programs would tend to score lower. In other words, the results can identify quality elements that are likely to suppress ratings in a block system in which all elements must meet the criteria for a given rating level.

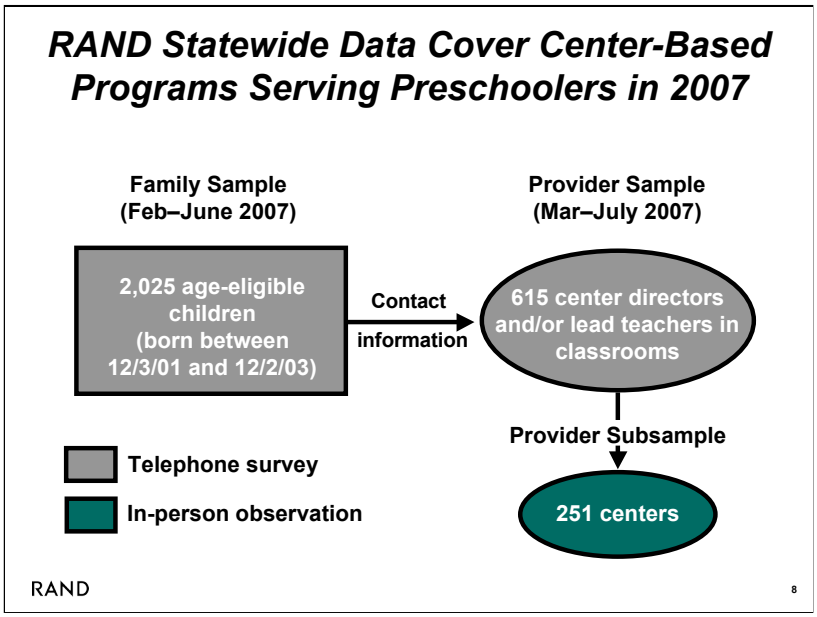
Ultimately, the goal is for these analyses to provide the California State Advisory Council on Early Learning (commonly known as the Early Learning Advisory Council, or ELAC) and other stakeholders with information about the consequences of decisions regarding the quality metrics to include in the QRIS and their associated cut points. This output also could help designers and implementers understand where TA is likely to be most needed once the QRIS is implemented. The data may further suggest QRIS design refinements or changes that might be made to avoid situations in which otherwise-qualified programs are prevented from moving to the next system tier. Finally, the analyses can inform future efforts in California to pilot the QRIS design and implementation.



This work relies on two sources of data with information on the ECE program features that map onto at least one of the five proposed quality elements in the QRIS. The first source of data is the provider sample in the 2007 RAND California Preschool Study. As discussed next, these data—collected through a telephone survey of lead classroom teachers and onsite observation of ECE classrooms—are representative of center-based ECE programs in California serving children one or two years prior to kindergarten entry (i.e., three- and four-year-olds).

The other source of data is the ratings of ECE providers participating in San Francisco County's GTQ initiative. The GTQ administrative data cover both center-based programs and FCCHs rated from June 2008 to early 2011. The children served by GTQ providers include infants, toddlers, and preschoolers.

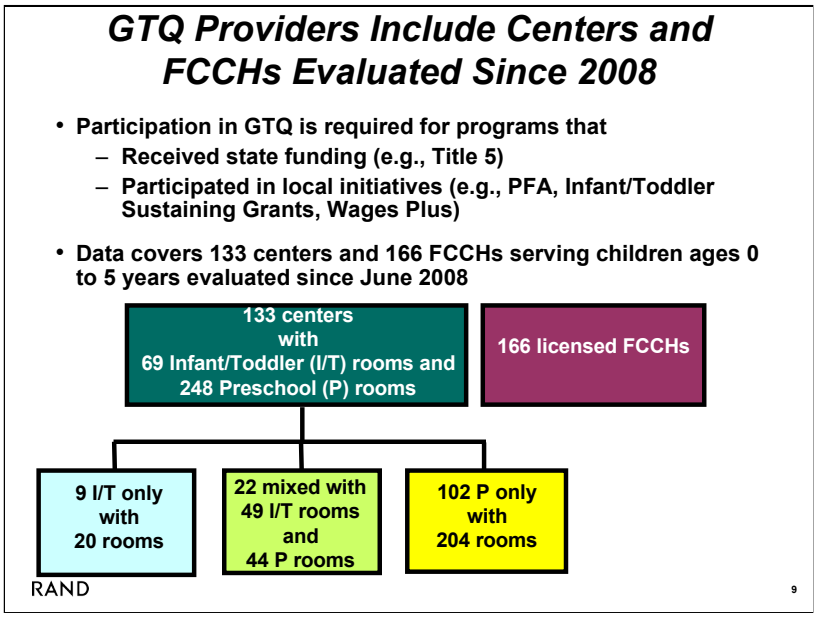
Following the study objectives, these two sources of data were used first to determine the distribution of single quality elements in the QRIS and then the distribution of multiple quality elements taken together, the latter to replicate the QRIS as closely as possible. The data were used to determine the expected share of programs that would fall into each tier of the rating system given specified cut points. The analyses also examined the relationships among different quality elements and pinpointed the elements that are likely to suppress a program's rating in the block system that underlies the proposed CAEL QIS design.



As noted, the RAND statewide data capture quality in center-based ECE programs serving preschool-age children as of 2007. These data were collected as part of the RAND California Preschool Study, a multiyear study of the adequacy and efficiency of preschool education in California (Karoly, 2009).

Most relevant for this study is that one component of the California Preschool Study designed and fielded a new survey to collect current information about the nonparental care arrangements experienced by California children before kindergarten entry and the quality of the care they receive (see Karoly et al., 2008, for further detail). As shown in this schematic, the data-collection effort began in early 2007 with a telephone survey of a representative sample of just over 2,000 California families with children in two kindergarten-entry cohorts (those who were age-eligible to enter in the fall of 2008 and those who were eligible to enter one year later). Parents with children in center-based settings were asked for permission to contact the child's provider. Telephone interviews were conducted with the director or lead teacher in the child's classroom or both for 615 public and private centers. Finally, specially trained observers collected objective information on program quality through onsite visits to a subset of 251 center classrooms in the provider sample.

This analysis relies on information collected through both the telephone interviews and onsite observations collected for the provider sample. That information includes onsite observations of group size, child-adult ratio, and the learning environment, as well as teacher reports of their education and training collected during the telephone interview. The results presented in this analysis are weighted to be representative of California center-based providers serving preschool-age children. (The results presented in Karoly et al., 2008, were weighted to be representative of preschool-age children in the state.) Missing data were imputed using methods described more fully in Karoly et al. (2008).

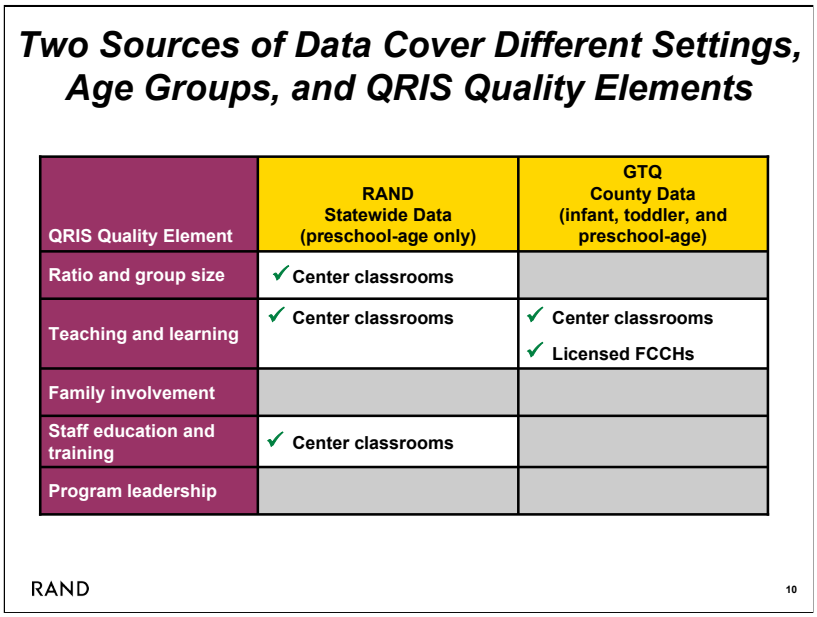


San Francisco's GTQ is a countywide QI initiative that began as a pilot program in 2002 (Gateway to Quality, undated). Programs that receive support from state- or county-funded programs are evaluated using the relevant ERS. The resulting ERS score forms the basis for a QI plan and TA. All of the participating programs receive support from one or more state or local programs, such as state Title 5 child development contract funding, the county's local Preschool for All (PFA) initiative, the Children's Council Infant/Toddler Sustaining Grant Program, and the Human Services Agency of San Francisco's Wages Plus program (a program that supplements the salaries of ECE classroom staff). Most of the providers have received some form of quality support, such as onsite coaching, training, or TA.

GTQ collects quality information on a three-year cycle. The programs we analyzed were rated between June 2008 and early 2011. Unlike the statewide data, the GTQ data cover both center-based providers and FCCHs serving children from birth to kindergarten entry. Within centers, all classrooms are evaluated.

As shown in the chart, the 133 centers in the administrative data file we analyzed have a total of 69 infant/toddler rooms and 248 preschool rooms. Only nine centers serve infants and toddlers exclusively, while 102 centers serve only preschool-age children. The remaining 22 centers serve all three age groups. A total of 166 licensed FCCHs are also in the data.

Because GTQ is a targeted system, the GTQ data are not necessarily representative of all providers in the county. The system is estimated to cover about half of licensed centers in the county and about one-third of licensed FCCHs (private communication from Gretchen Ames, project coordinator, GTQ, May 17, 2011). In the results presented later, the classroom is the unit of analysis.

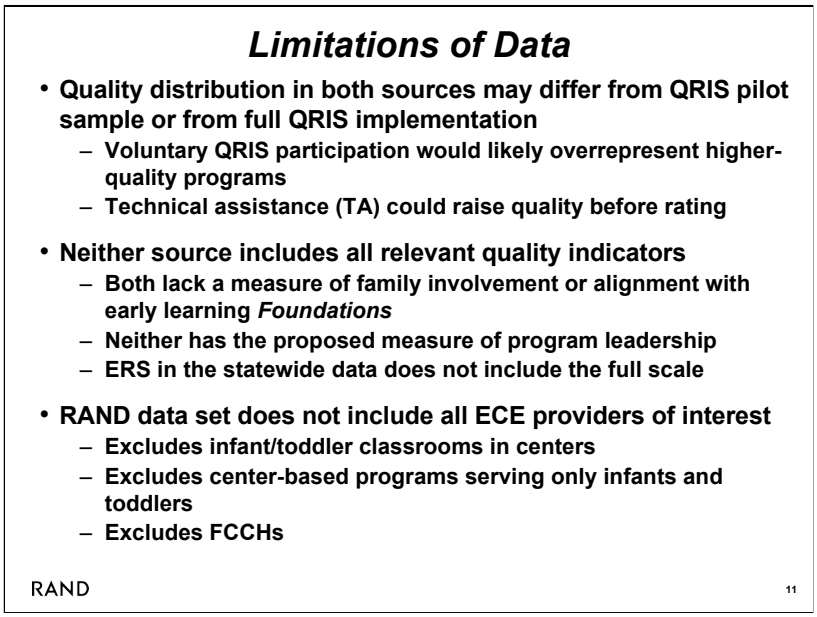


To fully analyze the proposed QRIS, the ideal data source would include information on all five quality elements for a representative sample of center-based providers and FCCHs serving infants, toddlers, and preschool-age children. The two sources of data available to us provide much, but not all, of this information.

As shown in the chart, the RAND statewide data capture three of the five quality elements for a representative sample of providers, but the data cover only center-based providers and only classrooms serving preschool-age children.

The GTQ county data cover both centers and FCCHs and all relevant child ages—infants, toddlers, and preschool-age children—but only one quality element is captured in the administrative data. In addition, as noted earlier, the GTQ data are limited to one county. Moreover, because GTQ serves a targeted group of programs, the providers in GTQ may not be fully representative of all providers in the county.

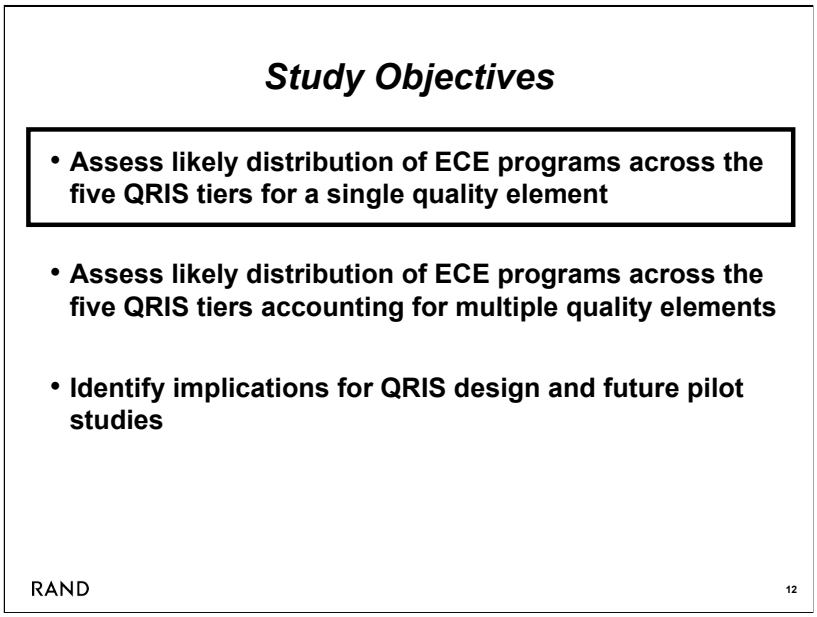
Despite these limitations, the two data sources provide information about the range of provider types that are expected to be covered by the proposed QRIS and allow us to examine key dimensions of quality that are likely to be included in California's rating system.



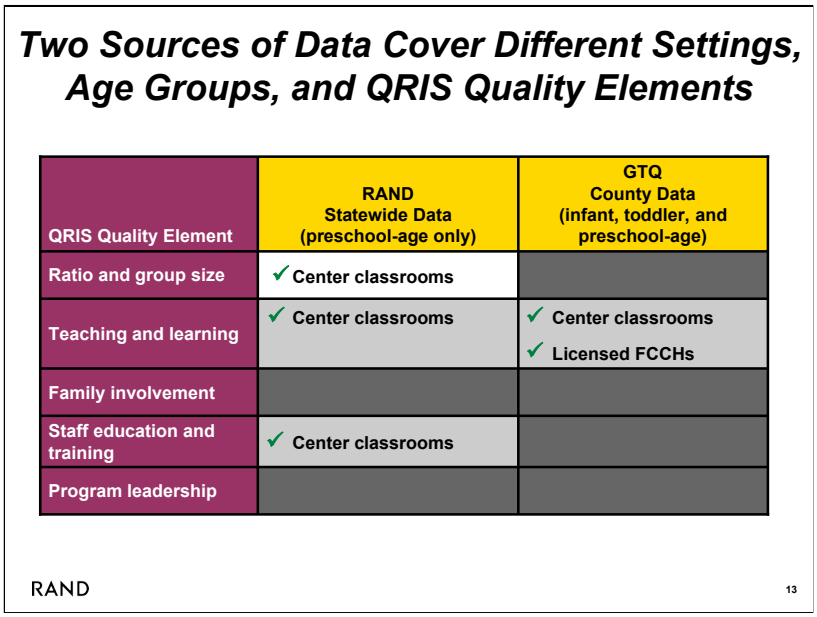
Given the features of the two data sources, it is important to keep in mind their limitations for the analyses we conduct. First, the composition of the providers captured in both data sources and their associated quality may differ from what might be expected at the start-up of a statewide QRIS or once a QRIS is fully implemented. In the case of the RAND statewide data, the representative sample includes both lower-quality providers that would be less likely to participate in a voluntary rating system, and higher-quality providers that would be more likely to participate. Thus, the representative sample captured in the RAND data may demonstrate a lower level of quality on average than what might be expected in the set of providers that would volunteer for a QRIS. In addition, if a provider receives TA prior to its first rating, such support would be expected to raise its rating. The distribution of quality in the GTQ data, albeit for one county, may reflect the selectivity of the providers in a rating system that would target providers receiving public subsidies, as well as the effect of TA on rated quality.

Second, both data sources lack information on the full set of quality indicators included in the advisory committee's proposed QRIS. There are no measures of family involvement in either source, nor are there indicators of alignment with the California infant/toddler or preschool learning Foundations (California Department of Education, 2008, 2009). Although some data on program leaders were collected in the statewide data (i.e., the center director's education background), there are no data on the proposed measure of program leadership. In addition, as discussed in more detail later, the ERS measure in the statewide data is a subset of the full scale. Given these limitations, we are not able to fully model the proposed rating system with the available data.

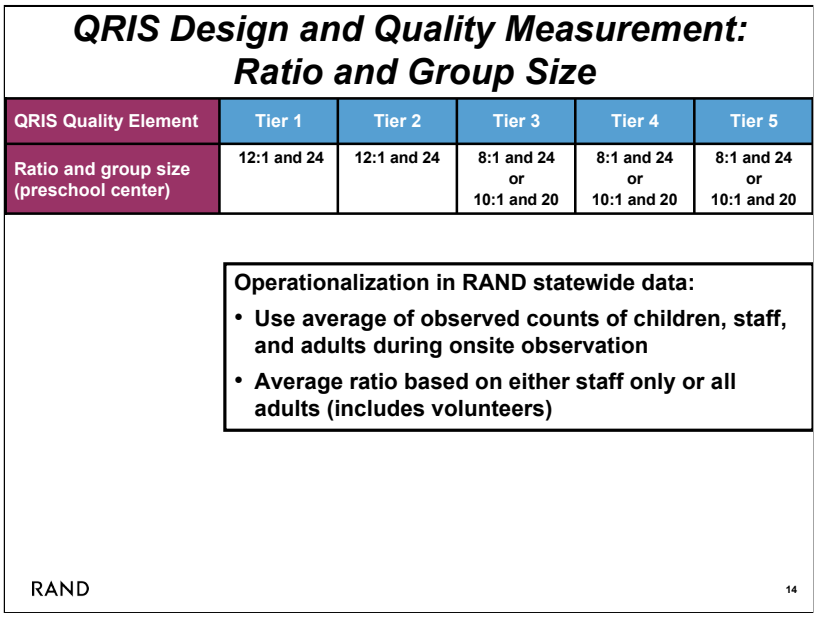
Finally, as noted previously, the RAND statewide data are limited to center-based programs that serve preschool-age children; no FCCHs are included. There is no information about infant/toddler classrooms in sampled centers. Nor are centers that serve only infants and toddlers included in the sample.



We now turn to the first set of results, in which we examine the distribution of programs across the rating tiers according to three of the five quality elements in the proposed QRIS. As we consider each quality element separately, we replicate as closely as possible the proposed measure of the element, and we employ the recommended cut points that define each tier in the rating scale. This analysis serves to highlight the quality elements for which programs score relatively high in the proposed rating tiers and for which they score relatively low.



We focus first on the ratio and group size quality element. Only the statewide data can be used to examine this quality component, which means that the analysis will pertain to center-based programs serving preschool-age children.



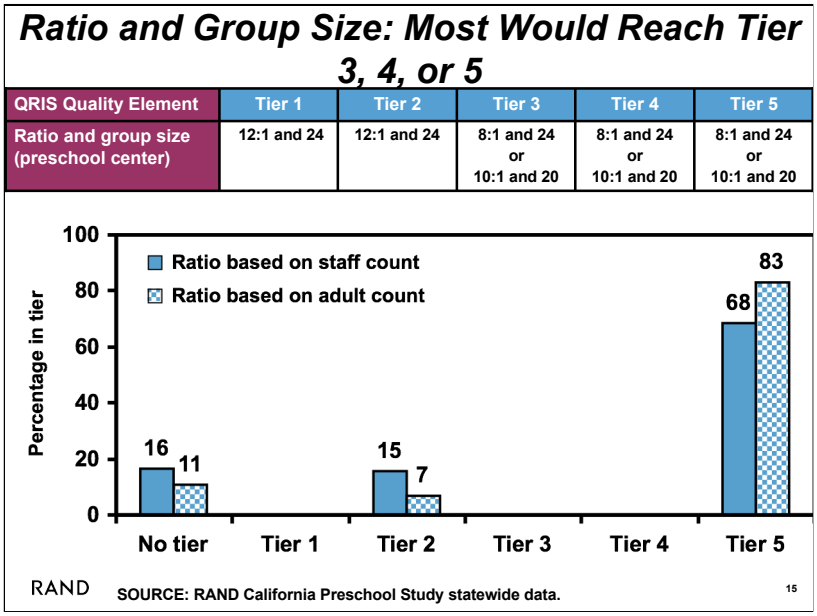
In the proposed QRIS, the advisory committee defined cut points for the child-adult ratio and the group size specific to the ages of children served (CAEL QIS Advisory Committee, 2010b). The top panel of this chart shows the cut points that would apply to center-based programs serving preschool-age children.[1] As shown, the first two tiers would require a ratio of children to adults that is no higher than 12 to 1 and a group size no greater than 24 children. The standard for Tier 3 or higher allows for two combinations of the ratio and group size: either a maximum ratio of 8 to 1 with no more than 24 children or a maximum ratio of 10 to 1 and no more than 20 children.

Our analysis of the RAND statewide data replicates the proposed ratio and group size standards for the center-based classrooms in the sample. Both indicators are based on the counts of children, staff, and other adults (e.g., volunteers) in the room made by the trained assessors during the observation period. Because both the ratio and group size were assessed about every 30 minutes during the observation period, we use the average value across all observations.[2]

Because the CAEL QIS Advisory Committee was not explicit as to whether or not volunteers could be counted when assessing ratios, ratios were calculated in two ways: In one approach, the ratio included only classroom staff; in the second approach, all adults in the classroom at the time that the ratio was assessed (including volunteers) were included in the count of adults. As reported in Karoly et al. (2008), on average, a center-based program in this sample serving preschool-age children had about 2.5 paid staff and one volunteer. Thus, we would expect that counting all adults, rather than just paid staff, when assessing ratios would allow more programs to meet a given ratio requirement.

[1]Preschool-age children are defined as age 31 months and higher for Tiers 1 and 2 and age 37 months and higher for Tiers 3 and above. A different set of cut points for the ratio and group size applies to infants in center-based programs, toddlers in center-based programs, and children in FCCHs.

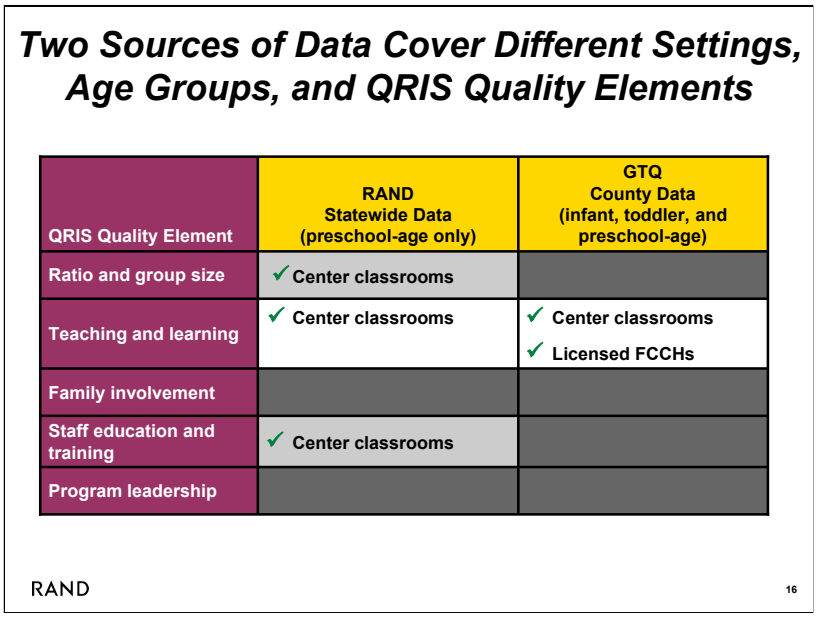
[2]On average, 6.6 counts were made during the observation period. See Karoly et al. (2008) for detail.



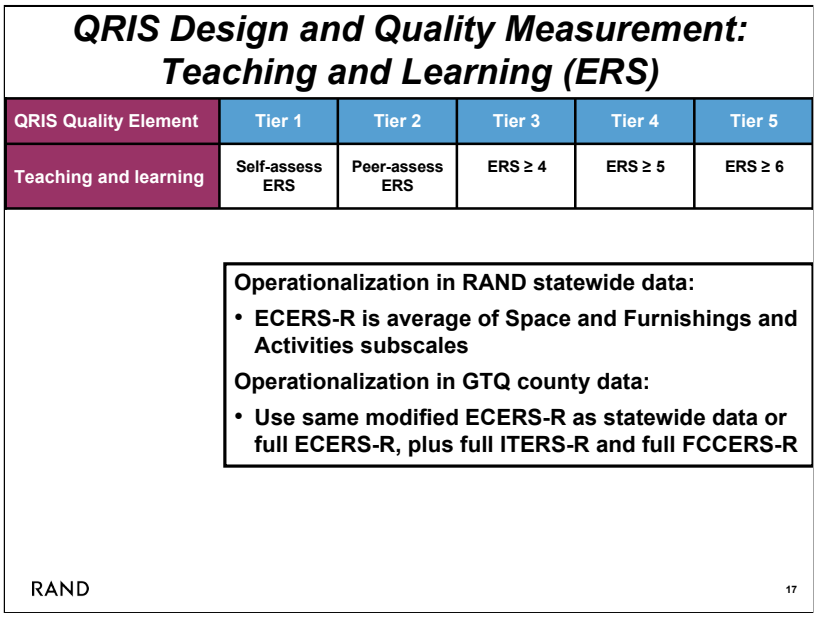
Ratio and group size data indicate that most center-based preschool programs would fall in Tier 3, 4, or 5. Because cut points for Tiers 3, 4, and 5 are the same, reaching Tier 3 would enable a program to also reach Tier 4 or 5, depending on its tier score on the other QRIS elements. Likewise, reaching Tier 1 also qualifies for Tier 2 status. Thus, graphically, we show programs in the highest tier for which their outcome on this element could qualify: either Tier 2 or Tier 5.

Using a ratio that counts only staff, 16 percent of programs fell below Tier 1 (labeled “no tier” in the figure), meaning that their average ratio exceeded 12 children to each staff member or the average group size exceeded 24 children. About 15 percent met the criteria to reach as high as Tier 2, while 68 percent could reach as high as Tier 5. As expected, when all adults (i.e., staff and volunteers) were counted in generating the ratio, more programs met the standard, thereby increasing the share of programs that would qualify for the higher tiers to 83 percent.

These results are consistent with the analysis of these data by Karoly et al. (2008), which showed that most children in center-based preschool programs in California are in programs that would meet nationally recognized benchmarks for the ratio and group size (e.g., the benchmarks set by NAEYC and NIEER that were used by the advisory committee to establish the standards for the higher tiers).



We now turn to the second quality element; this one can be examined using both sources of data. This means we can assess how different types of providers—center-based programs serving children of different ages, as well as licensed FCCHs—would be rated under the proposed system. Because both data sources cover center-based classrooms with preschool-age children, we can determine how the distribution of quality might vary if all providers were in the system (as seen in the statewide representative sample) versus only those providers that receive public funding or supports (as seen in the GTQ data).



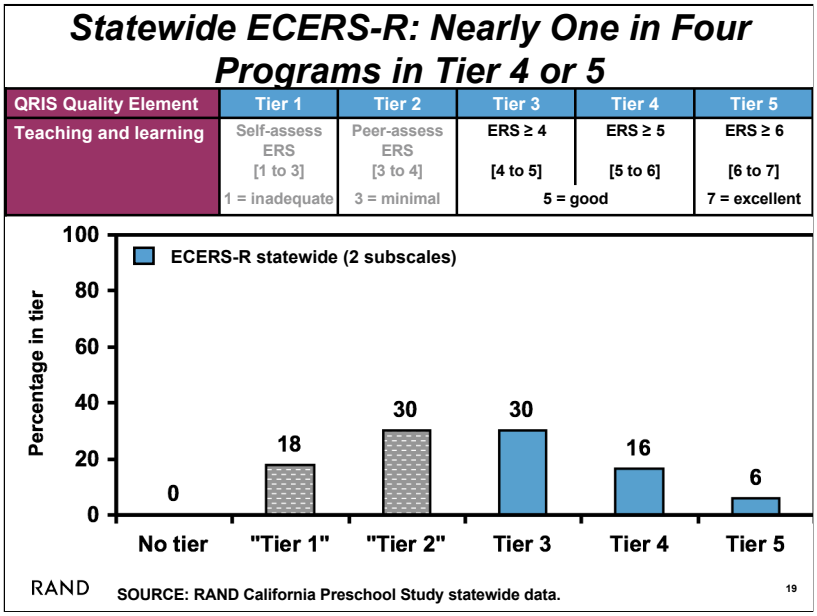
The teaching and learning quality element was designed to capture aspects of program structure (beyond ratio and group size), as well as aspects of teacher-child interactions, that are expected to affect the quality of the learning environment. As proposed by the advisory committee, this element would rate center- and home-based providers based on the appropriate ERS (CAEL QIS Advisory Committee, 2010b). (At the two highest tiers, the teaching and learning element also involves an assessment using the Classroom Assessment Scoring System [CLASS]. We'll take up this measure after the ERS analysis. As proposed, this element also assesses a program's alignment with the state's early learning standards, a feature we are not able to assess with our data.)

The advisory committee recommended using the family of environment rating tools developed by researchers at the Frank Porter Graham Child Development Institute at the University of North Carolina at Chapel Hill, which are widely used in research and in other QRISs (Tout et al., 2010). For center-based programs, the latest version of the tools consists of the 39-item Infant/Toddler Environment Rating Scale–Revised (ITERS-R) used in rooms with children up to age 2.5 years (Harms, Cryer, and Clifford, 2006) and the 43-item Early Childhood Environment Rating Scale–Revised (ECERS-R) used in rooms with children ages 2 to 5 years (Harms, Cryer, and Clifford, 2005). The 38-item Family Child Care Environment Rating Scale–Revised (FCCERS-R) is used in home-based settings (Harms, Clifford, and Cryer, 2007). Each of these multi-item scales results in an overall classroom score that can range from 1 to 7. The developers of the ERSs equate a score of 1 with "inadequate" care quality, a score of 3 with "minimal" quality, a score of 5 with "good" quality, and a score of 7 with "excellent" quality.

As shown in the chart, the two lowest tiers would require either a self-assessment or peer assessment, respectively, using the relevant ERS; no minimum score is specified. For the next three tiers, the ERS score would be assessed by a trained,

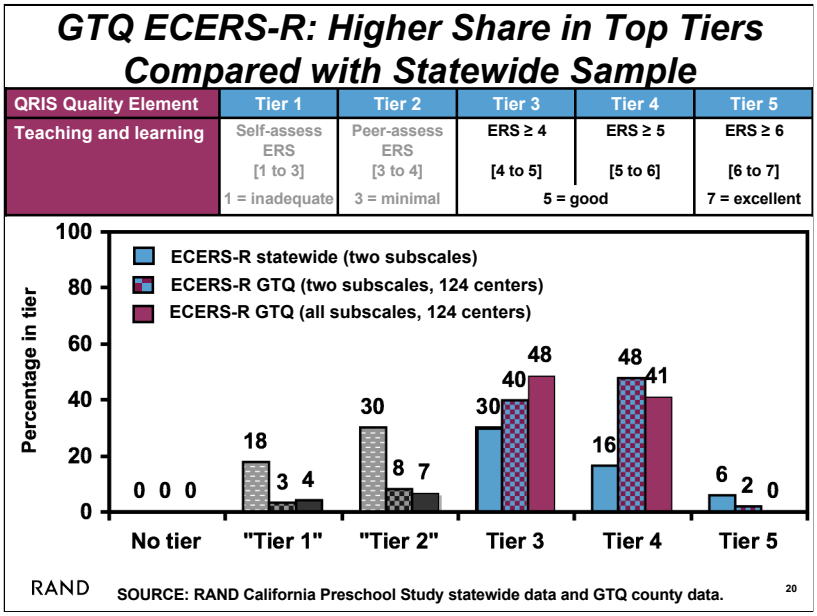
independent observer. A successively higher minimum score is required for successive tiers; a score of 6 or higher out of 7 total is required to be rated in the top tier.

In both the RAND and GTQ data, ERS scores were determined by trained, independent assessors, making data collection consistent with the requirement of independent ERS measurement for Tier 3 or higher. Because the RAND statewide data provide observations of classrooms with preschool-age children, the ECERS-R is the available measure, but only two of seven subscales were collected: Space and Furnishings (eight items) and Activities (ten items) (see Karoly et al., 2008, for further detail). The GTQ administrative data contain ITERS-R or ECERS-R scores based on the ages of the children in the center-based classrooms, and the FCCERS-R for the licensed home-based programs in the system. Because we had access to the individual ECERS-R items, we constructed the modified ECERS-R scale employed in the RAND data in the GTQ data set and use it in some of our analyses.



This chart shows the results for the analysis of the modified ECERS-R using the statewide data. For this sample of preschool classrooms, the analysis revealed that almost one-quarter of the programs (22 percent) would be classified into Tier 4 or 5 on this element as they scored a 5 or higher. Another 30 percent had ECERS-R scores that would award them Tier 3 status.

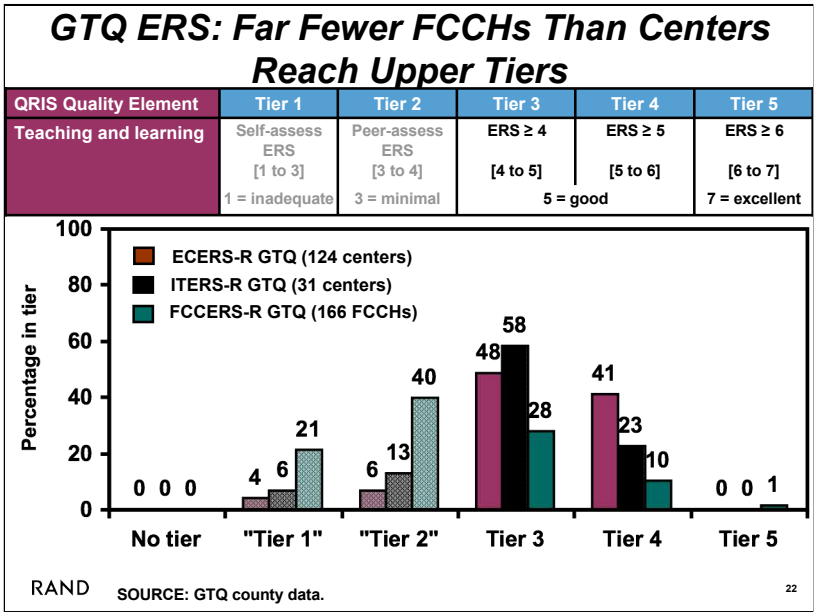
As noted earlier, the advisory committee proposed that the ECERS-R score would not be used to assign programs to the two lowest tiers. However, for those programs that would not score high enough to reach Tier 3, we assigned them to two “pseudo tiers” (shown in gray, patterned bars) to determine how far away their scores would be from the Tier 3 cut point. As shown in the chart, the “Tier 2” group is defined as those up to one scale point away from the Tier 3 criteria, while the “Tier 1” group is anywhere from one to three scale points away from the Tier 3 level. This classification shows that about three in five of the programs that fall below Tier 3 are at most one scale point away, while about two in five have a larger shortfall. The descriptive labels attached to the seven-point ERS scale would classify the “Tier 1” group as providing care of “inadequate” quality. The “Tier 2” group would be viewed as providing care of “minimal” quality. Because no programs score below a one on the ERS, our definition of the lower two tiers does not leave any programs below “Tier 1.”



This chart takes the results in the prior chart (first set of bars) and adds to them the distribution of center-based classrooms serving preschool-age children in GTQ using two versions of the ECERS-R. The second set of bars (patterned) shows the distribution of classrooms in GTQ that results when the same two ECERS-R subscales measured in the statewide data (i.e., Space and Furnishings and Activities) are used in place of the full ECERS-R scale. Because this contrast uses the same version of the modified ECERS-R, the resulting differences in the distribution of providers can be attributed to (1) any differences in provider quality in San Francisco county versus the state as a whole and (2) any differences in provider quality based on which providers participate in GTQ. Regardless of the explanation, the chart shows that a considerably higher share of GTQ providers would be classified in Tiers 3, 4, and 5 (90 percent for GTQ versus 52 percent statewide). Most of the difference occurs in the share in Tier 3 (40 versus 30 percent) and Tier 4 (48 versus 16 percent). According to both data sources, only a very small fraction of providers would be classified in Tier 5 and the share is actually lower among the set of GTQ providers (2 percent versus 6 percent statewide). The higher scores on the ECERS-R for the GTQ programs than for the statewide sample is consistent with findings in Karoly et al. (2008) that various quality elements were somewhat higher among publicly funded programs in California (e.g., Head Start and Title 5) than for providers without public subsidies.

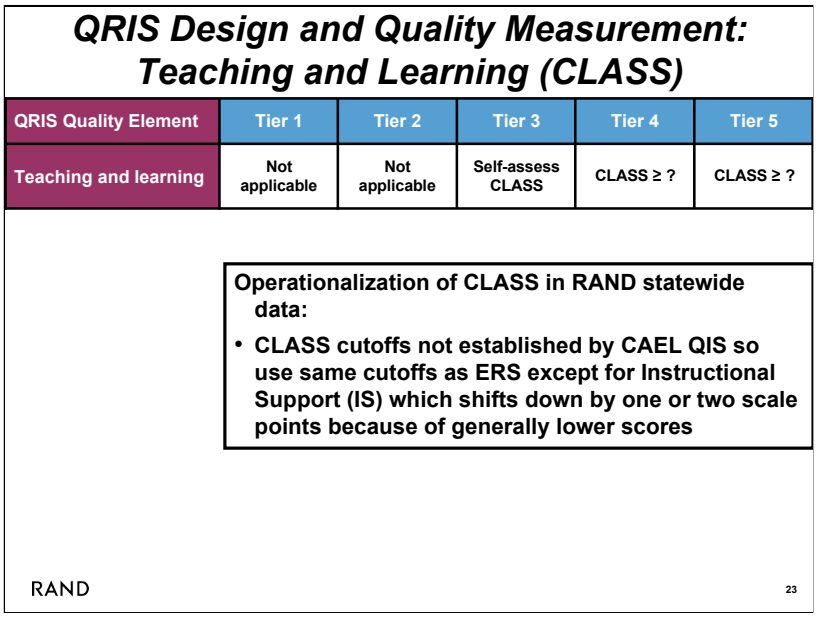
The third set of bars (darkest shading) shows the classification of providers using the full ECERS-R for the GTQ preschool classrooms. Contrasting these results with the second set of bars shows the effect, for the same GTQ set of providers, of using the modified ECERS-R to rate programs versus the full ECERS-R. As seen in the figure, the differences are concentrated in Tiers 3 and 4, with the Tier 3 share increasing by 8 percentage points when the full ECERS-R is used, while the Tier 4 share falls by 7 percentage points. The share in Tier 5 dips from 2 to 0 percent. These results show

that using the modified ECERS-R shifts the ratings of preschool classrooms in center-based programs from Tier 3 into Tiers 4 and 5, which indicates that programs score better on the Space and Furnishings and Activities subscales of the ECERS-R. If this pattern also holds for the statewide sample, it suggests that the share of programs evidenced in the first set of bars for the statewide data overestimates the share that would be classified in Tier 4 and underestimates the share that would be in Tier 3.



The GTQ data also enable examination of how center-based classrooms with infants and toddlers and FCCHs would be classified based on the appropriate ERS. The first set of bars in this chart replicates the results in the prior chart for GTQ center-based preschool classrooms based on the full ECERS-R scale. The center-based infant/toddler rooms, classified using the ITERS-R, are shown in the second set of bars, while the third set of bars describes FCCHs classified using the FCCERS-R.

Overall, this comparison indicates that preschool classrooms would be most likely to reach Tier 3 or higher (89 percent), while the FCCHs would be the least likely to achieve that Tier (38 percent). The percentage of infant/toddler rooms that would reach Tier 3 or higher (81 percent) would be closer to that of the preschool rooms. Notably, none of the three setting types would find any appreciable share of classrooms in Tier 5, the highest tier. Tier 4 would be reached by just over four in ten preschool rooms, slightly more than two in ten infant/toddler rooms, and just one in ten FCCHs. Upward of 20 percent of FCCHs would fall into the lowest pseudo tier, indicating an inadequate level of quality.

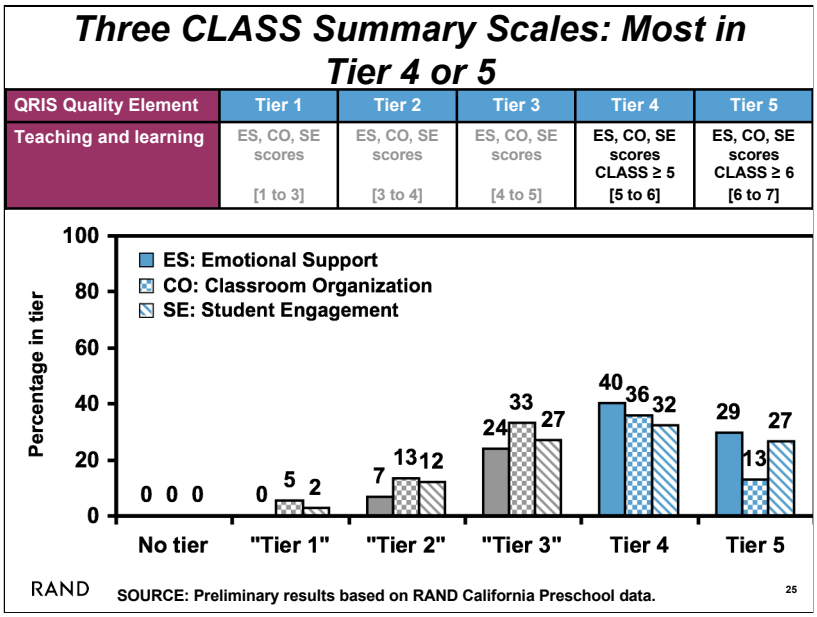


As noted earlier, the advisory committee also proposed that the teaching and learning element incorporate the CLASS at higher tiers (CAEL QIS Advisory Committee, 2010b). The preschool version of the CLASS assesses the quality of teacher-child interactions in ECE settings, with ratings on a seven-point scale in three domains: emotional support (ES), classroom organization (CO), and instructional support (IS) (Pianta, La Paro, and Hamre, 2008). The CLASS was collected as part of the statewide RAND study but is not available in the GTQ administrative data. The version of the CLASS used in the RAND study included a fourth domain: student engagement (SE) (see Karoly et al., 2008, for details).

The committee determined that the CLASS would first be introduced as a tool for program self-assessment at Tier 3. By Tier 4, the CLASS would be collected by independent assessors and used as an additional criterion, but no specific cut point values were recommended by the QRIS designers. For purposes of our analysis, we assumed that, at Tiers 4 and 5, the same cut points used for other ERSs would apply to the CLASS, i.e., a 5 or above in Tier 4 and a 6 or above in Tier 5.

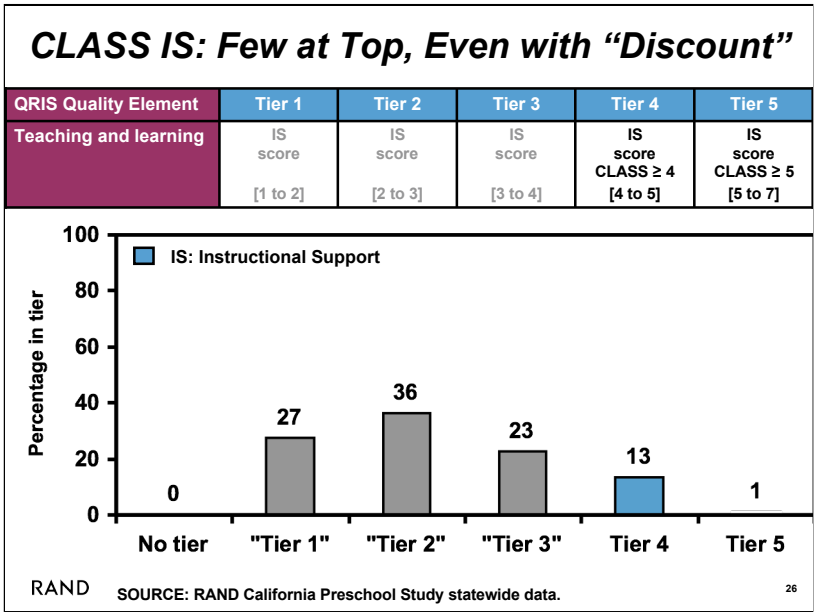
One issue with CLASS is that, unlike ECERS-R, ITERS-R, or FCCERS-R, subscale scores usually are not combined to produce a single summary score. Moreover, subscale scores tend not to be the same. In particular, programs typically score much lower on the IS scale: often two to three points lower. So, when constructing a composite score, we required, at Tier 4, a 5 or higher for the ES, CO, and SE subscales but only a 3 or higher on IS. We assumed that each of these cut points would be set one scale point higher to reach Tier 5. The lower values for the IS cut points are consistent with data from Oklahoma, which suggests that it is possible to positively affect child outcomes with CLASS IS scores that are at a lower level than the other CLASS subscales (Gormley and Gayer, 2005; Gormley et al., 2005; Phillips, Gormley, and Lowenstein, 2009). Because CLASS subscale scores are usually not

combined to produce a single summary score, we report subscale scores separately but also report a composite score to model how a QRIS might use the CLASS in a block system.



The CLASS analysis revealed that most of the programs in the statewide sample attained Tier 4 or 5 status on the ES, CO, and SE subscales. Indeed, almost 70 percent of programs attained Tier 4 or 5 status on the ES subscale; the comparable figure for the CO subscale was 49 percent; the figure for the SE subscale was 59 percent. More programs fell into Tier 4 than Tier 5; the difference across Tiers 4 and 5 is greatest for the CO subscale (36 percent in Tier 4 versus 13 percent in Tier 5) and least for the SE one (32 versus 27 percent).

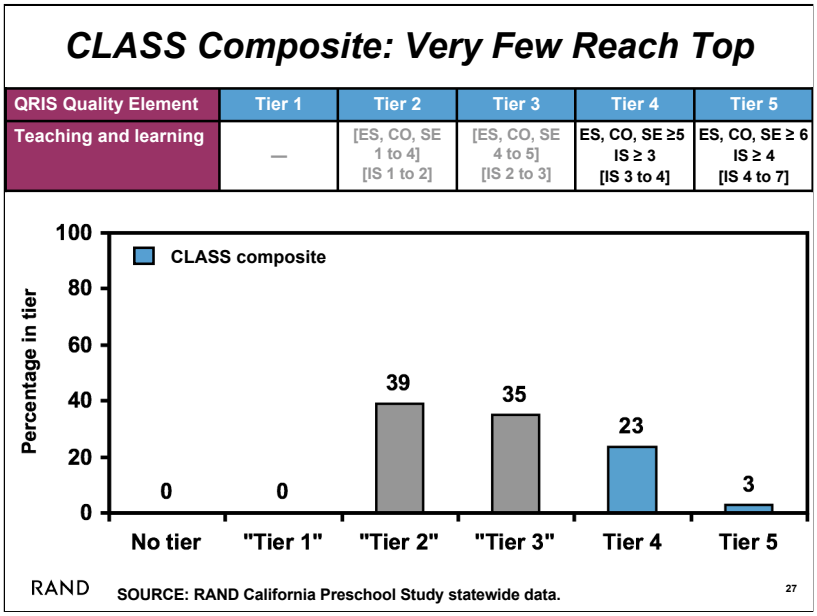
Again, we defined pseudo tiers (shaded in gray) for scores that fell below the lowest specified cut point. In this case, the cut point falls by one scale point with each successively lower tier. Most of the programs that would not achieve Tier 4 or 5 status would be classified into Tier 3 because they achieved subscale scores of at least a 4 but below a 5. The percentage of programs in Tier 3 were 24, 33, and 27 percent, respectively, for the three CLASS subscales. Of the programs in Tiers 1 and 2, the highest percentages were found on the CO subscale, in which 18 percent received scores that classified them into the two lowest tiers.



As we had anticipated, program scores on the CLASS IS subscale were far lower than they were for the previous three subscales. In this chart, we have allowed for a one-point "discount" on the IS subscale, in which Tier 4 could be achieved with a score of 4 or higher and Tier 5 with a score of 5 or higher. Even with this lower cut point, only 13 percent of programs would be classified into Tier 4, and effectively none would reach Tier 5.

Of the remaining three lower pseudo tiers, most of the programs would achieve what we have defined as Tier 2 status (36 percent); 23 percent of programs would be classified into Tier 3 and 27 percent into Tier 1.

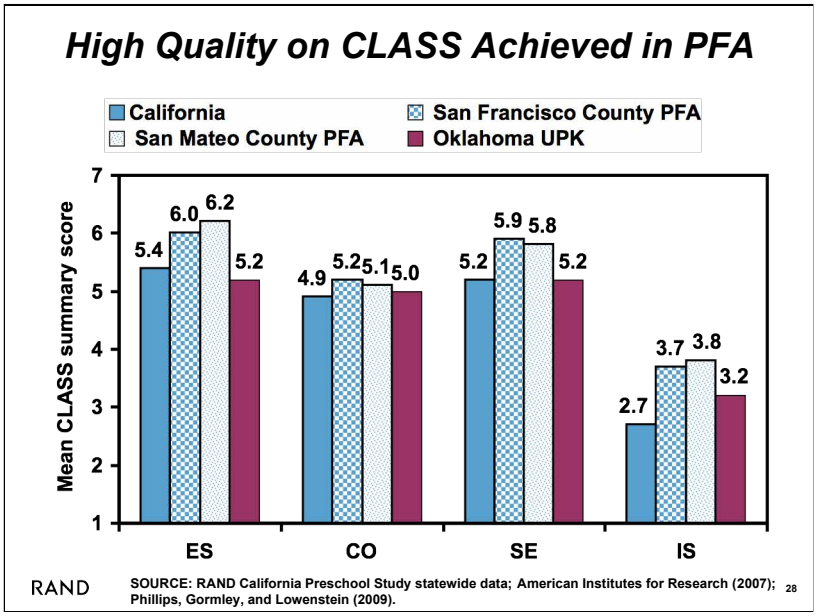
It is important, in considering the results for the IS subscale, to compare what appear to be low scores in the sample of California center classrooms with those found in other studies. Most notably, despite the fact that the average CLASS IS subscale score for a sample of programs in the Tulsa district was just 3.2 (Phillips, Gormley, and Lowenstein, 2009), the Oklahoma universal prekindergarten program (UPK) has been evaluated and shown to produce substantial learning gains for young children (Gormley and Gayer, 2005; Gormley et al., 2005). This average score is equivalent to our pseudo "Tier 3" level. Given these findings, when we create our CLASS composite measure next, we further "discount" the cut point required for the IS subscale by two scale points relative to the cut point specified for the other three CLASS subscales.



Despite the fact that CLASS subscales are usually not combined and there was no recommendation concerning how to combine them from the CAEL QIS Advisory Committee, we constructed a composite measure that accounts for scores on each of the four subscales to model how a QRIS might use the CLASS in a block system. As shown in the chart, to reach the highest tier, we required the ES, CO, and SE subscales to each reach a score of 6 or higher, the same cut point used for the ERS and shown in the earlier slide for these CLASS domains. But because of lower IS scores, we allowed that subscale to be as low as a 4 (two scale points lower) and still qualify a program for the highest tier (the cut point that defined Tier 4 for the IS in the prior chart). The cut points drop by one scale point for each lower tier, but the IS score cut point is always allowed to be two points lower than the other three subscales.

When combined in this way and using the cut points we have set, few programs would be classified in the top tiers. About one-quarter (23 percent) of programs would attain a Tier 4 level, but only 3 percent would reach Tier 5 status. Almost four in ten programs would attain Tier 2 status (39 percent), and just over one-third would reach Tier 3 (35 percent).

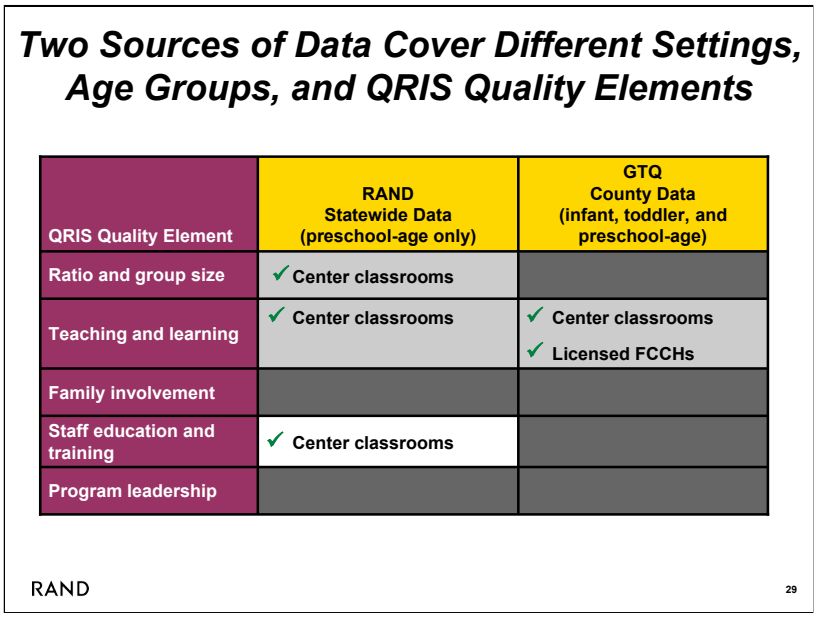
Given the subscale data presented earlier, it is clear that the low scores on the IS subscale are constraining programs from attaining a higher tier, even when we define the cut point to be two scale points lower for that subscale.



To achieve more perspective on the CLASS scores that we reported for the statewide data on preschool classrooms, this figure plots average CLASS subscale scores for the statewide RAND study sample, as well as results for three quality-focused preschool programs. The second and third bars show average CLASS subscale scores for classrooms in the PFA initiatives in San Mateo and San Francisco counties (American Institutes for Research, 2007). The fourth bars show the average scores for Oklahoma's UPK program that serves four-year-olds, based on a sample of classrooms in the Tulsa school district (Phillips, Gormley, and Lowenstein, 2009). Both the PFA initiatives and the Oklahoma program include high standards for program features that are associated with high quality and better child developmental outcomes.

Overall, compared with the statewide sample, these results show that higher scores on the CLASS subscales, including the IS subscale, can be achieved in center-based preschool programs that invest in quality. At the same time, IS subscale scores are consistently as much as two points below scores attained for the other subscales, even in programs that are considered to be of high quality. Compared with the California statewide sample in the first bar, the two PFA programs score higher on each of the subscales, with more-pronounced differences (between 0.7 and 1.1 scale points) for the ES, SE, and IS subscales.The statewide scores are similar to those for Oklahoma's UPK program for all but the IS score, for which the Oklahoma score tops the statewide score by 0.5 points. Given the narrower range in the IS scores, the size of that gap is a meaningful difference.

The results for California's statewide sample—representative of center-based programs serving three- and four-year-olds—reflect important differences from the three other programs. Although participation in PFA and Oklahoma's UPK requires programs to meet strict quality criteria and engage in a QI process designed to improve ECE quality, the California programs were not selected because of their involvement in a QI initiative (although some may have been involved in QI efforts).



We now turn to the third quality element: staff education and training. As with the ratio and group size measures, this element is available only in the statewide data, so the results will pertain only to center-based programs serving preschool-age children.

QRIS Design and Quality Measurement: Staff Education and Training

QRIS Quality Element	Tier 1	Tier 2	Tier 3	Tier 4	Tier 5
Staff education and training (center)	12 units ECE 6 mos. exp 21 hrs PD	24 units ECE 1 yr exp 21 hrs PD	24 units ECE + 16 units GE 2 yrs exp 21 hrs PD	AA in ECE or with 24 units ECE 2 yrs exp 21 hrs PD	BA in ECE with 48 units ECE 2 yrs exp 21 hrs PD

Operationalization in RAND statewide data:

- **Use measures of staff education, training, and permit status to approximate design**

QRIS Quality Element	Tier 1	Tier 2	Tier 3	Tier 4	Tier 5
Staff education and training (center)	Assoc. Teacher Permit or CDA	24 units ECE	Teacher Permit	AA in ECE or Master Teacher Permit	BA in ECE

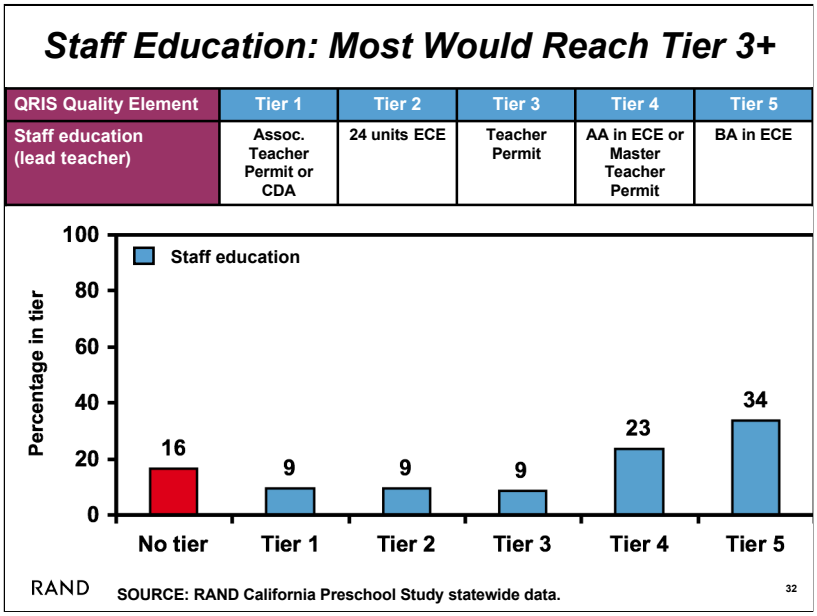
RAND 30

The advisory committee recommended a set of standards for the staff education and training element that incorporates postsecondary coursework and degrees, experience in the field, and ongoing professional development (PD), with some differences for staff in center-based programs versus FCCHs in the first two tiers (CAEL QIS Advisory Committee, 2010b). The top portion of this chart shows the requirements specific to center-based staff, presumably the lead classroom teacher.

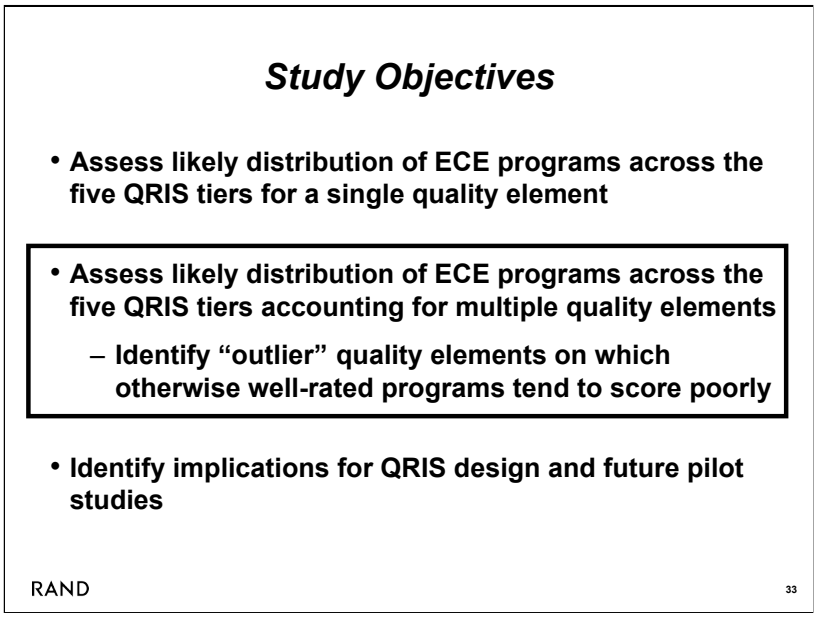
The information available in the RAND statewide data does not allow us to include all of the requirements proposed for each tier. Notably, we focus on capturing the education requirement as closely as possible; we do not incorporate the experience or PD requirements. In all likelihood, this means that our analysis may overstate the number of programs that would be in the higher tiers because some classroom staff may meet the education criteria for a given tier but not the criteria pertaining to experience or PD.

In terms of education level, the advisory committee recommended 12 units of ECE education for Tier 1. Because the RAND data did not record the actual number of ECE units for the lead teacher, we proxy this standard by requiring the teacher to have a California child development Associate Teacher permit (which also requires 12 ECE units) or a Child Development Associate (CDA) credential (information that was obtained in the survey). This is a more stringent requirement, however, because some classroom teachers will have that number of credits but will not have applied for the permit. The RAND data did record whether the lead teacher had at least 24 units of ECE credits, so that requirement is applied for Tier 2. The child development permit system is also used to proxy the education requirement in Tier 3. Specifically, the Teacher permit requires 24 units of ECE credits and 16 general education (GE) units, a level that is equivalent to the Tier 3 requirement. But again, not all staff who may have met that requirement will necessarily hold the associated permit.

For Tier 4, the advisory committee recommended an associate degree (AA) in ECE or 24 semester units of ECE coursework. To approximate this requirement, we define Tier 4 as having either an AA degree or a Master Teacher permit, which can be obtained with 24 units of ECE or a bachelor's degree (BA) or higher. Finally, for Tier 5, the advisory committee specified that the BA in ECE must include 48 units of ECE coursework, while the BA in ECE in our Tier 5 criteria does not specify a particular number of ECE units.



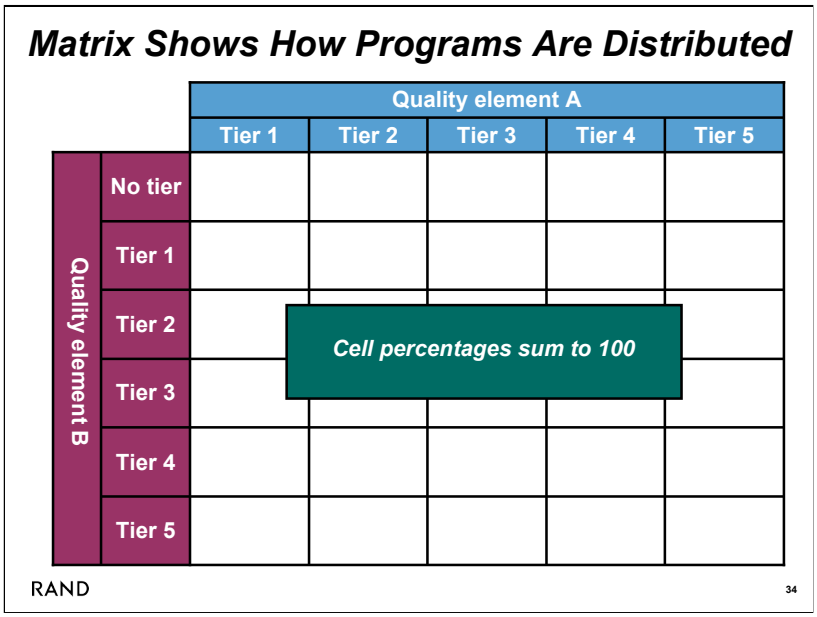
When we apply these criteria to define the tiers, the RAND statewide data indicate that the lead teacher's education credentials would allow most center-based preschool programs to reach Tier 3 or above. Indeed, two-thirds of programs (66 percent) would be classified into Tier 3, 4, or 5, with half of those programs (34 percent) reaching Tier 5 status. Fewer than one in five programs (18 percent) would find themselves in Tier 1 or 2. Notably, the remaining 16 percent of classrooms have a lead teacher who, according to the RAND data, would not meet even the Tier 1 education standard. Again, this may overstate the share of programs in this group because not all ECE staff with 12 or more ECE credits necessarily apply for the Associate Teacher permit, nor do they have a CDA. Thus, some unknown fraction of the no-tier group would probably rate in the first tier.



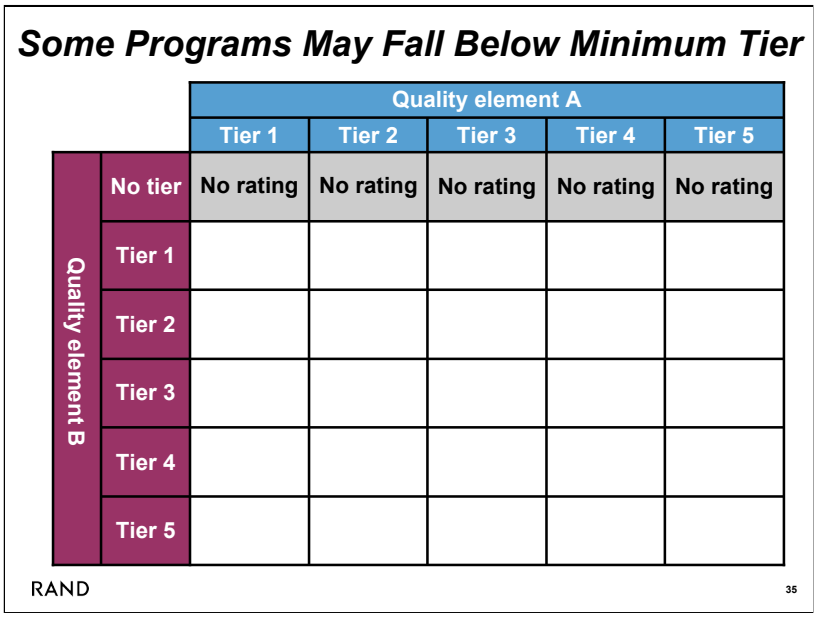
A second goal of our analysis was to examine multiple quality elements together as a way of identifying "outlier" elements on which otherwise well-rated programs score poorly. The identification of such "outlier" elements is important in a block system design because such elements may prevent programs from being rewarded for improvements by moving up a tier in their summary rating.

At the design phase of a QRIS, information about the existence and performance of "outlier" elements may help designers decide whether to modify the cut points for these elements, as we did with the CLASS IS discount in our analyses described above. Alternatively, if cut points are not altered, this information may identify areas in which TA and QI efforts need to focus so that outlier elements do not keep programs from attaining higher summary ratings. This latter approach may be more feasible in already-functioning QRISs, in which changes to cut points may be difficult to make.

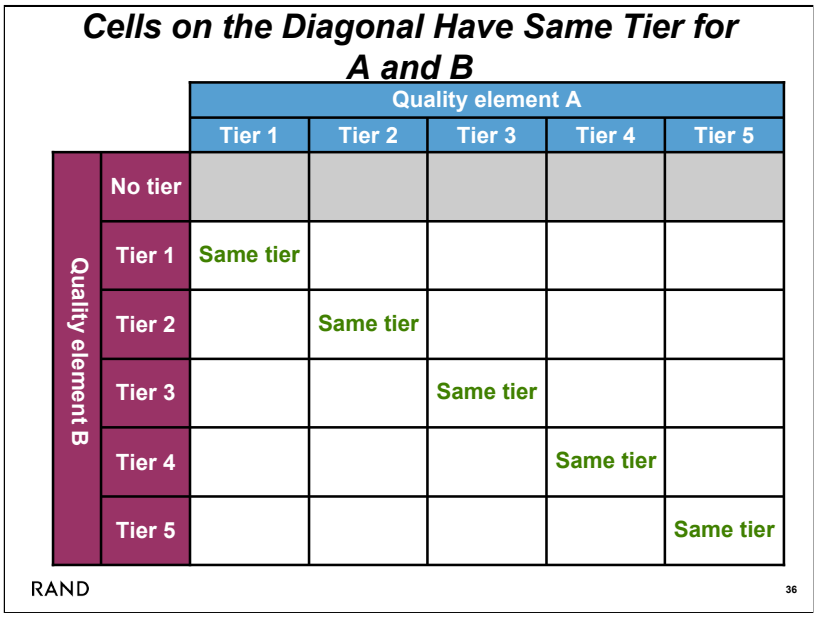
Our analysis of combinations of quality elements is limited to the statewide data on center-based preschool programs because the GTQ county data contain information on only one quality element. Given that our statewide data on preschool center-based classrooms capture three of the five proposed quality elements, we start by examining pairs of QRIS elements, an approach that is also easier to portray visually, as we do in the slides that follow. This analysis will reinforce the expectation that few programs will reach the higher tiers when we account for measures of the learning environment (e.g., ERS or CLASS), regardless of what other quality element we account for at the same time. For this reason, we do not reinforce the point by considering all three available quality elements simultaneously because even fewer center-based programs would meet all three elements at the highest tiers.



The matrix presented on this slide is used in subsequent slides to indicate where programs fall on the two quality elements that we consider together. With columns representing quality element A and rows representing quality element B, cell entries represent the percentage of programs that fall at a given tier level on the two elements under consideration. As noted, cell entries sum to 100 percent, indicating that all possible combinations of the two quality elements are included in the matrix.

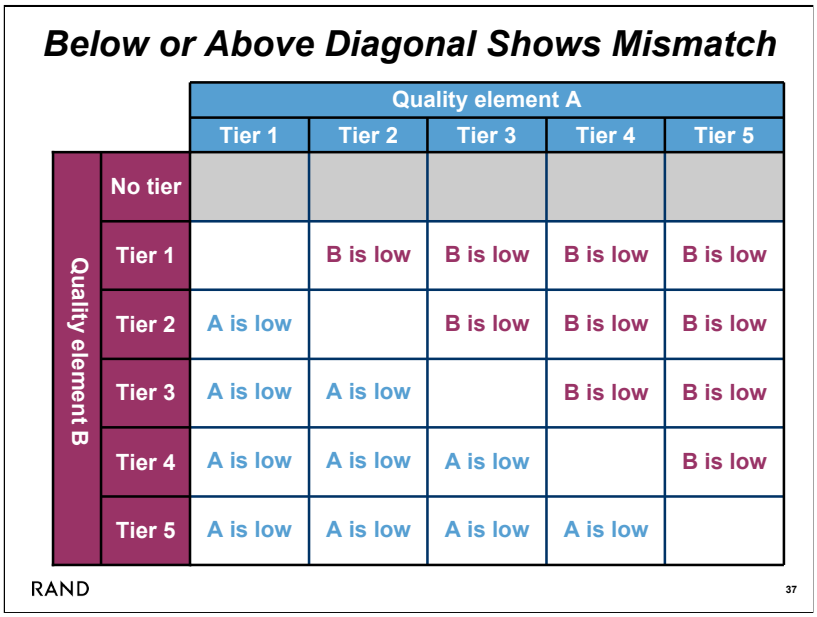


As noted on this slide, some programs may score below the Tier 1 criteria on one or both of the elements displayed on a given slide. Those programs will be located somewhere in the first row (or first column) and labeled in that row (or column) as the no-tier group. In all the combinations we view, only one quality element at a time has a no-tier group. That quality element is always listed in the matrix rows.



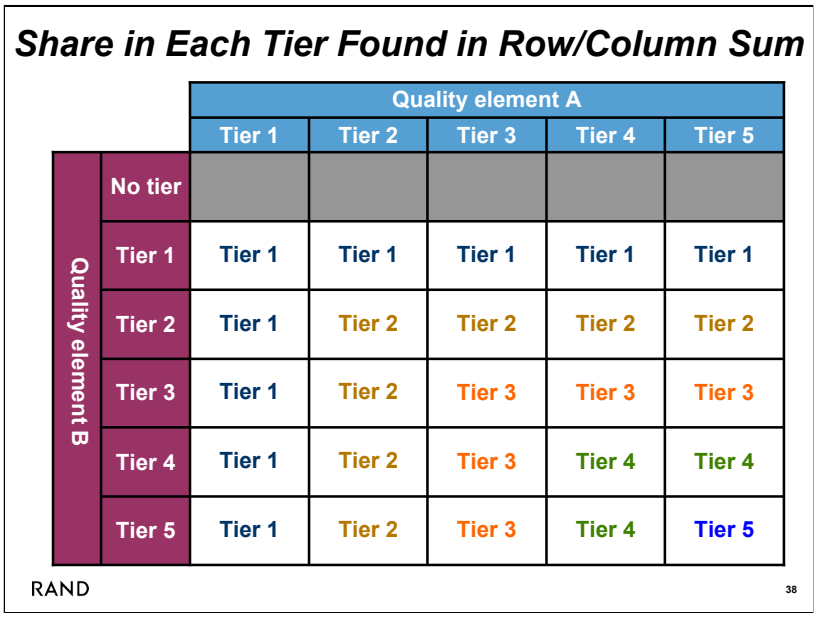
Cells that make up the diagonal of the matrix indicate that a program has scored in the same tier on both of the elements under consideration. Scoring in the same tier on all measured elements is an underlying goal of block system designs. By not permitting programs to attain a higher summary rating until all elements meet requirements for the next tier, block systems encourage programs to seek the diagonal placement shown on this matrix.

In contrast, QRISs designed around point systems allow programs to compensate for low scores on one or more elements with high scores on others. This is possible because summary ratings are based on the total number of points that a program achieves; a program may score high on some elements but low on others and still receive sufficient points to improve its summary rating.

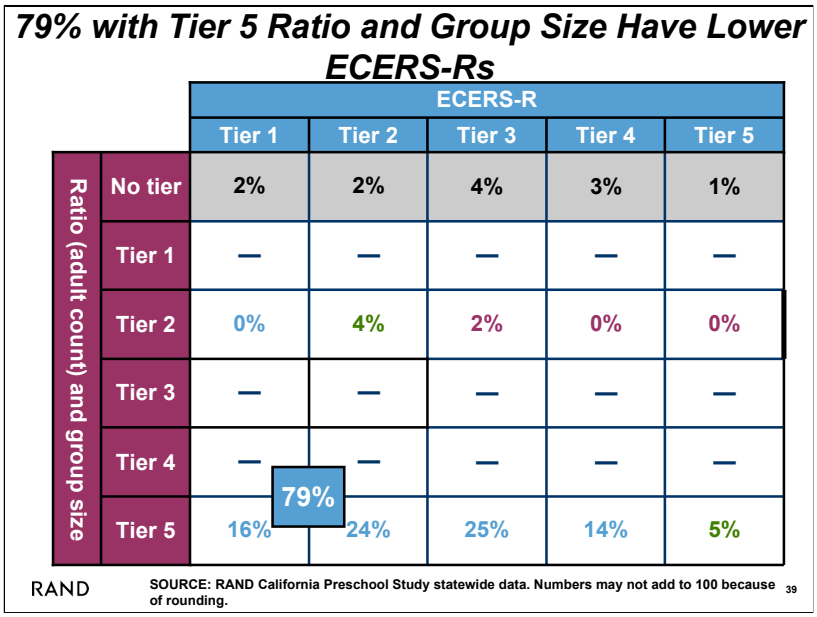


Off-diagonal placement indicates that a program scored in a higher tier on one of the paired QRIS elements than in the other paired element. As represented here, cells that are above the diagonal (upper right corner) are ones for which program scores on quality element B put the program in a lower tier than those for quality element A. For cells below the diagonal, the reverse pattern holds. Whether above or below the diagonal, the lower quality element would prevent a program from attaining a higher summary rating because all elements have to meet the criteria for a given tier. The higher tier achieved on one element may focus a program on improving those elements that currently do not meet criteria for higher tiers.

In contrast, in point systems, differences in tier levels across QRIS elements do not necessarily constrain the summary rating. Indeed, a higher score on one or more elements, if sufficiently high, might be able to pull a program with lower scores on other elements into a higher summary rating.



Finally, we note in this chart that the share of programs in any given tier based on rating both quality elements together can be found by summing the relevant diagonal and off-diagonal cells. The share of programs meeting Tier 5, for example, is the single cell in the bottom right corner. The share in Tier 4 is the sum of the appropriate diagonal cell and the two off-diagonal cells. The Tier 3 share is the sum of five cells, and so on as shown.



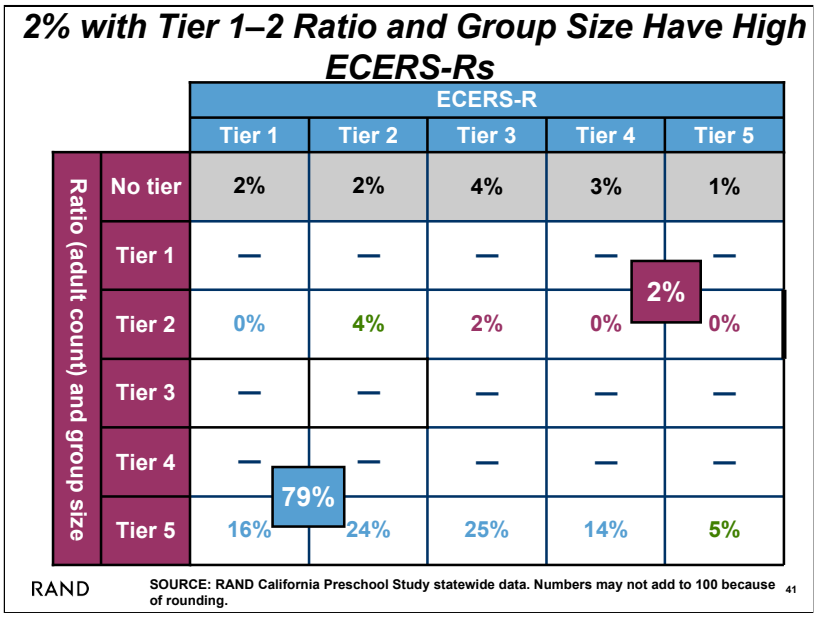
We start by considering the first two quality elements together: ratio (based on counting adults) and group size with teaching and learning, represented by the ECERS-R because the statewide data were collected in center-based preschool programs. The columns represent the ECERS-R tier that a program would attain; the rows represent the ratio and group size tier. Cell entries represent the percentage of programs that score at the specified tier level on each element. As noted earlier, cell entries sum to 100 percent (subject to rounding error), indicating that all relevant combinations are included in the matrix.

First, note that, in the case of the ratio and group size tiers, 12 percent of programs (the sum in the gray-shaded cells in the first row) would not achieve even Tier 1 based on their observed ratio or group size.[3] Second, as indicated earlier, the ratio and group size criteria are the same for Tiers 1 and 2 and the same for Tiers 3, 4, and 5. This means that the cells in the Tier 1, 3, and 4 rows are not applicable because we assume that programs could reach as high as Tier 2 or as high as Tier 5 based on their ratio and group size because those tiers are the maximum they can attain if they meet the specified thresholds. Third, the cell percentages indicate that just 5 percent of programs would reach Tier 5 accounting for both ratio and group size and teaching and learning, as captured by the ECERS-R (bottom right corner cell).

With this understanding, we can see that nearly four in five programs could reach Tier 5 based on their ratio and group size tier rating but would be constrained from achieving Tier 5 by a score on the ECERS-R that would place them in a lower tier. For about 40 percent of the programs, the ECERS-R score would place them in either

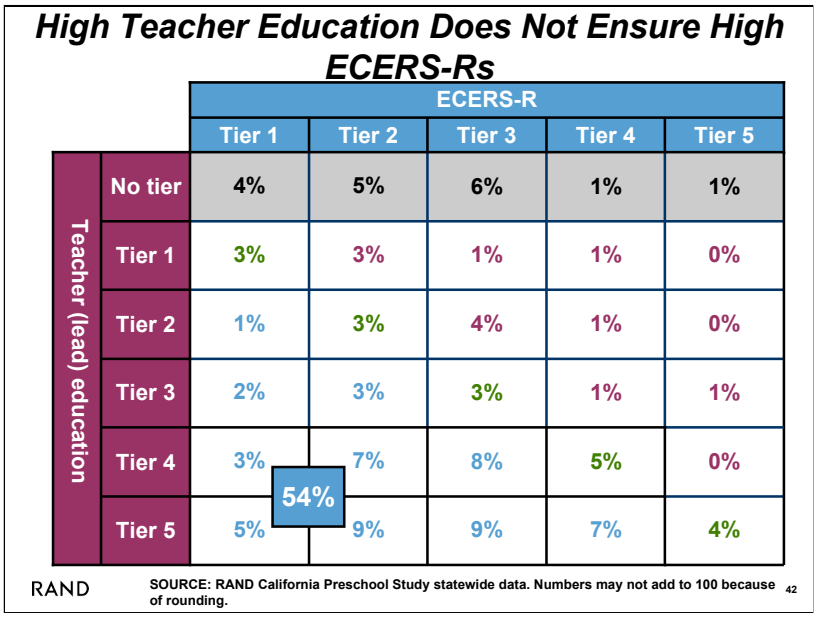
[3]Note that earlier, 11 percent fell in the no-tier group when the ratio was based on adults in the room. The 1-percentage-point difference shown here is because of rounding error.

Tier 1 or Tier 2, indicating that substantial improvement in their ECERS-R score would be required to move them to the same tier they achieved for the ratio and group size element. Another 39 percent would be in either Tier 3 or Tier 4, so the shortfall from the Tier 5 ECERS-R standard is smaller.



Looking at the same data, this slide shows that very few programs score in a higher tier on the ECERS-R than on ratio and group size (cells above the diagonal). Indeed, just 2 percent of programs had an ECERS-R rating in Tier 3 or above with a ratio and group size rating in a lower tier.

Together, these two slides indicate that programs would appear to be more likely—in the absence of further QI supports—to qualify for higher tier status on the ratio and group size element than on the teaching and learning element as captured by the ECERS-R. One explanation for this pattern is that the greater regulation around ratios and group sizes—whether through state Title 22 licensing requirements or the requirements associated with publicly funded programs, such as State Preschool—has compelled programs to meet standards on these indicators. In contrast, no standards exist regarding aspects of quality reflected in the ECERS-R, which includes a greater focus on the process of providing care and supporting early learning in the classroom than on the classroom's structural features. Indeed, the lack of standards around the more process-oriented features of quality is a key reason that those who want to improve ECE quality argue for the implementation of a QRIS.

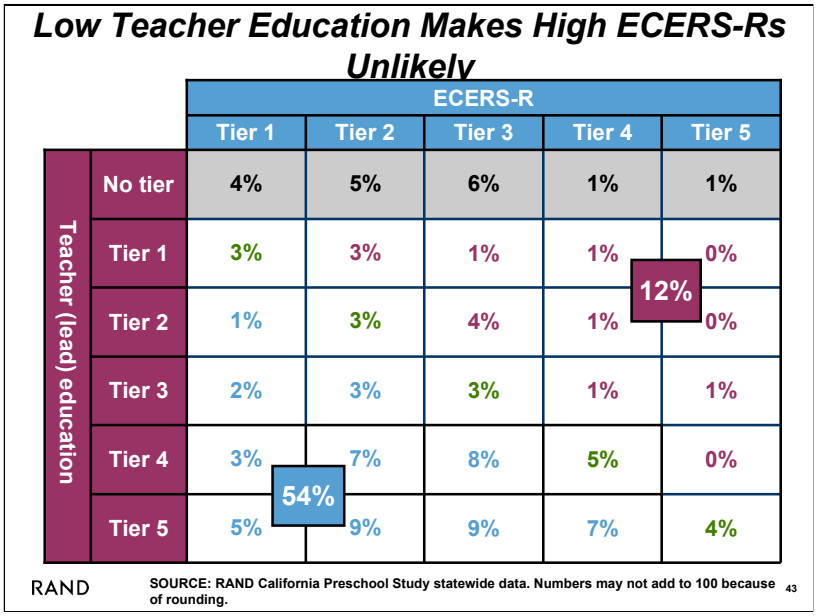


The matrix presented on this slide indicates the tier in which programs fall based on teacher education (rows) and ECERS-R score (columns, as before). Although the advisory committee was not specific regarding which staff member's education to count, we have used the lead teacher, as identified in the statewide data, as with our earlier analysis of this quality element. In that analysis, we found that the share of programs for which the lead teacher's education background would not reach Tier 1 status could be as high as 16 percent, the same percentage represented by the sum of the gray-shaded cells in the first row (subject to rounding error).

Overall, these results show that 18 percent of programs would score at the same tier level for both elements (the diagonal cells). Again, only a very small fraction (4 percent) would attain Tier 5 status on both the ECERS-R and teacher education.

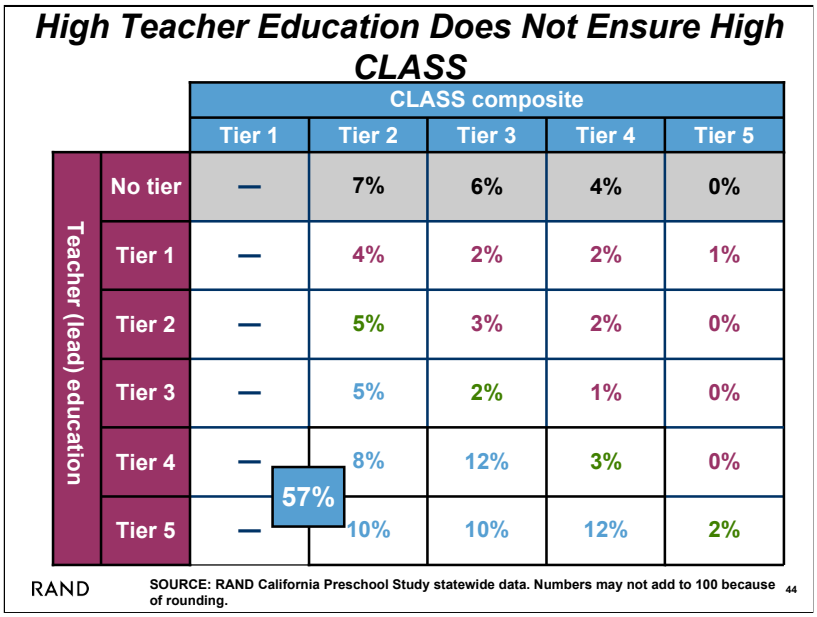
In nearly all cases of mismatched tiers (off-diagonal cells), programs would score higher on teacher education than on the ECERS-R. More than half of all programs (54 percent) would attain a tier rating based on lead teacher education that is higher than the tier they would attain based on ECERS-R score. Most of the programs in this situation (48 percent) are ones for which lead teacher education level would qualify the program for Tier 4 or 5 status, indicating that the teacher has an AA or BA degree in early childhood education.

Knowing that this analysis does not capture the experience and PD requirements associated with the staff education and training element in the proposed QRIS, we expect that some programs would qualify for lower tiers on the education element if those requirements were also accounted for. These findings nevertheless indicate that high levels of teacher education do not ensure a high ECERS-R score.



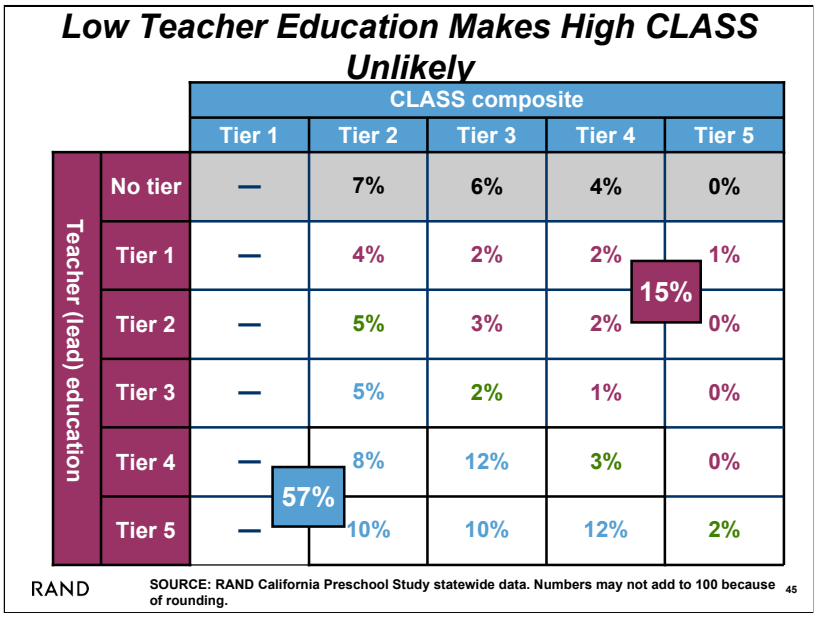
Further considering these same two quality elements, this slide highlights that relatively few programs (12 percent) would score in a higher tier on the ECERS-R than on lead teacher education. At the higher end, just 9 percent of programs would have an ECERS-R rating in Tier 3 or above with a teacher education rating in a lower tier.

In sum, these two slides demonstrate that programs would likely be able to qualify for higher tier status on the staff education and training element than on the teaching and learning element. This is an interesting finding given that teacher education is one of the most costly elements to fund and improve. It suggests, as noted above, that little attention has been focused on improving the aspects of ECE quality captured by the ECERS-R. The findings nevertheless suggest that high levels of teacher education may be important in promoting high-quality classroom environments. Although we see that a high level of teacher education does not ensure high ECERS-R scores, when classrooms are led by less educated teachers, the likelihood of scoring high on the ECERS-R is low.



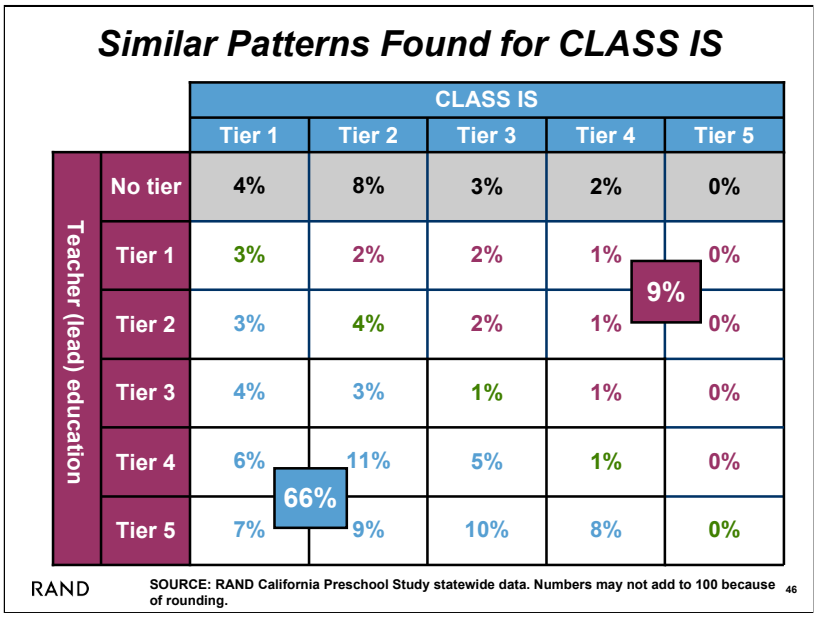
Thus far, we have examined the pairwise combinations of ratio and group size with ECERS-R and teacher education with ECERS-R and found that the ECERS-R is the quality element that would most constrain programs from achieving higher tiers on the proposed QRIS. Given that the CLASS is also under consideration for inclusion in the teaching and learning element, we examine how programs would be rated accounting for both teacher education and the CLASS. For purposes of considering the CLASS, we use the same composite measure presented earlier, in which the cut points for three of the four subscales match those for the ECERS-R at each tier but the IS subscale cut point is set at two scale points lower. Because the CLASS is proposed to enter the rating system only at Tier 4 and higher, the columns for Tier 2 and Tier 3 on the CLASS should be considered pseudo tiers. (As before, Tier 1 is not defined for the CLASS composite measure.)

The results shown in this matrix indicate that just 12 percent of programs would score at the same tier level for both elements and that only a very small number (2 percent) would achieve a Tier 5 rating on both quality elements. Just as with the ECERS-R, in nearly all cases of mismatch, programs score higher on teacher education than on the CLASS. Indeed, more than half of all programs in our sample (57 percent) would attain a tier rating based on lead teacher education that is higher than the tier they would attain based on their CLASS composite score. Mirroring the results for the ECERS-R, most of the programs in cells below the diagonal would be found in the last two rows, in which the lead teacher has either an AA or a BA, thereby qualifying the program for Tier 4 or 5. Once again, these findings indicate that high levels of teacher education do not guarantee a comparable rating on the teaching and learning element, this time captured by the CLASS score.



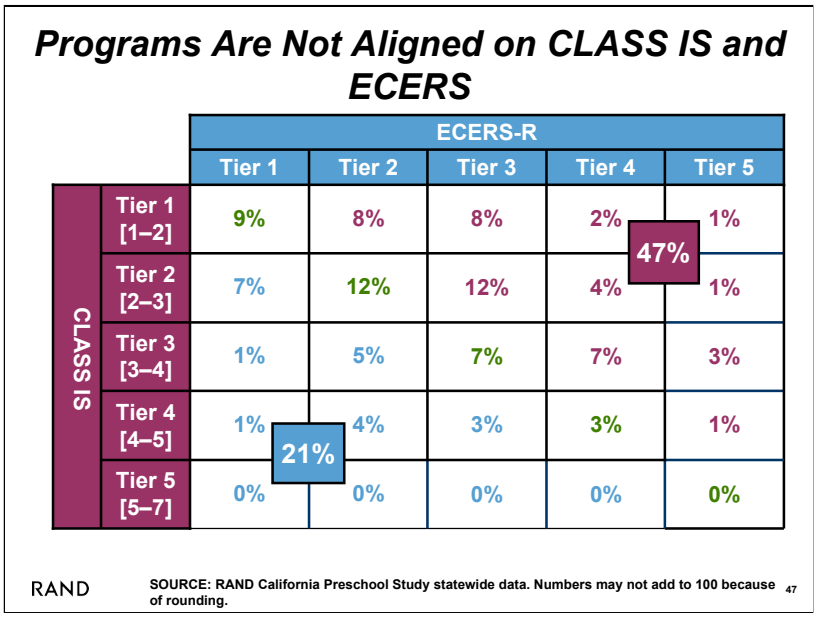
As before, when we consider the cells above the diagonal, we find that few programs would score in a higher tier on the CLASS composite score than on lead teacher education. In the statewide sample, 15 percent of programs would have a CLASS tier level that exceeded their teacher education tier level. Further, just 11 percent of programs would achieve a CLASS score in Tier 3 or above but a teacher education rating in a lower tier.

The findings for the CLASS echo those observed for the ECERS-R. Regardless of which measure we use to capture the teaching and learning dimension, our statewide data indicate that center-based programs serving preschool-age children, as currently configured, will be rated in a higher tier based on teacher education than they would be according to their rating on the classroom environment. Once again, not many programs achieve a CLASS score that would place them in Tier 3 or higher when the lead teacher has no postsecondary degree, but those programs with a lead teacher with an AA or BA do not consistently score at high levels on the CLASS.



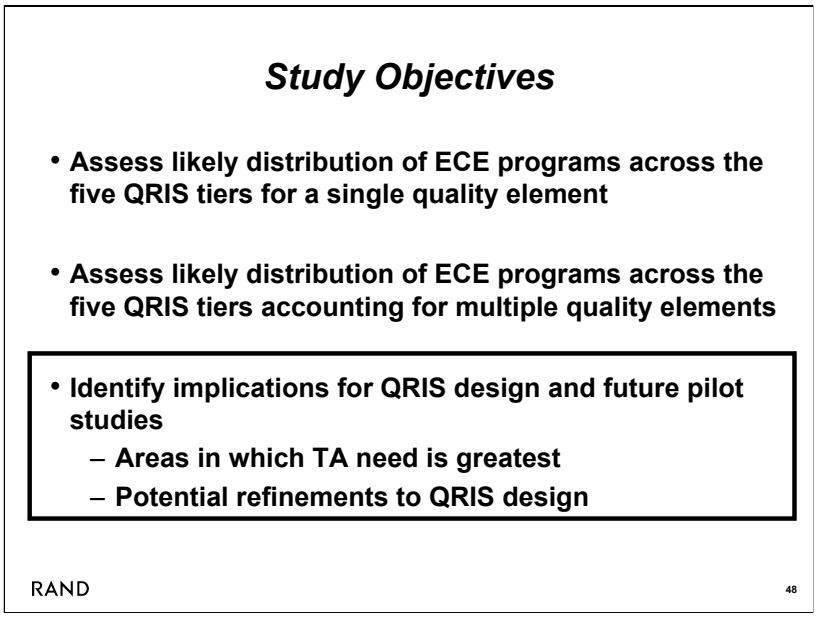
For the analysis presented in this matrix, we move away from the CLASS composite and once again place a spotlight on the IS subscale because it has been shown to be the CLASS subscale most closely associated with improved child cognitive outcomes (Hamre and Pianta, 2005; Howes et al., 2008). As in the earlier charts, we examine a measure of the teaching and learning environment in combination with teacher education.

These results based on the CLASS IS subscale reinforce what we have already seen using the CLASS composite and, before that, the ECERS-R. Even using the lower cut points for the IS subscale used earlier (two scale points lower than those used for the other CLASS subscales or the ECERS-R), only 9 percent of programs would reach the same tier on both teacher education and the IS subscale, as shown in the diagonal; no programs score at the Tier 5 level on both elements. An even higher fraction of programs would now fall below the diagonal (66 percent); most of these programs would qualify for Tier 4 or Tier 5 status based on level of teacher education alone but would not qualify because of lower CLASS IS score. Finally, only a small fraction of programs (9 percent) would achieve a higher IS tier based on CLASS IS score than they would on teacher education. Consistent with our findings for the ECERS-R and the CLASS composite is the fact that few programs achieve a CLASS IS score that would place them in Tier 3 or higher when the lead teacher has no postsecondary degree, but those programs with a lead teacher with a degree do not often score at high levels on the CLASS IS.

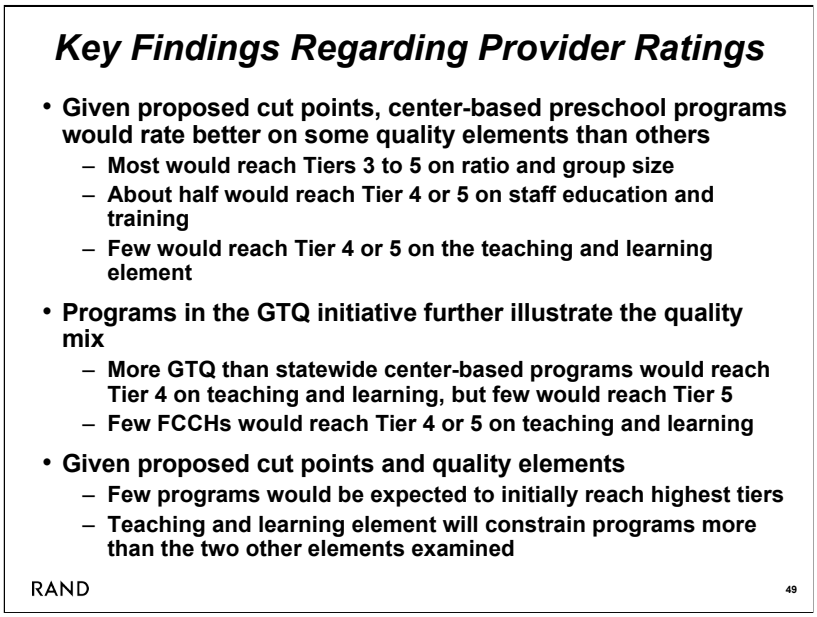


Given the similarity in the pairwise relationship between teacher education and the ECERS-R and the CLASS, we expected the two potential measures of teaching and learning to be highly correlated. The matrix presented on this slide illustrates that relationship, focusing on the IS subscale of the CLASS (shown in the rows) and the ECERS-R (shown in the columns).

In this pairwise comparison, we find that nearly one-third (31 percent) of programs would score at the same tier level for both elements, a relatively high level, as we expected. Even so, about one-fifth of all programs (21 percent) would attain a tier rating based on the CLASS IS score that is higher than the tier they would reach based on ECERS-R score. An even larger share (47 percent) would achieve a higher tier on the ECERS-R than on the IS subscale. When we use the cut points we have employed for the two quality components, the results together indicate that achieving a high tier level on the ECERS-R does not ensure that a program will score well on the CLASS IS subscale. However, the odds of scoring well on that subscale without a high tier rating on the ECERS-R are low.



A key objective for this work was to use the information captured in the data presented here to inform the decisions of the the California ELAC and other stakeholders regarding the design of a statewide QRIS and the types of issues that merit further investigation as part of a pilot effort. We believe that these analyses point to areas in which TA support is likely to be most needed and in which it could have the greatest impact. We also highlight some design issues that warrant further consideration. We begin by highlighting the key study findings and caveats that inform our discussion of the implications of this work.



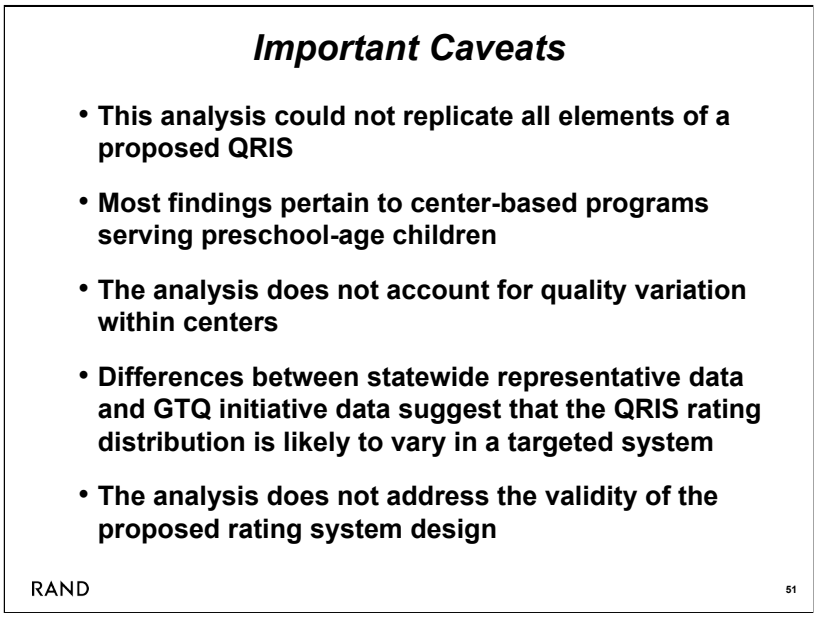
The analyses presented in this briefing support the key findings regarding provider ratings shown here. First, by separately examining three of the five quality elements proposed for California's QRIS, we find that most programs in our statewide center-based sample would rate better on some quality elements than others, a result that is likely to generalize to other settings. In particular, we found that 70 to 80 percent of programs would qualify for a Tier 3 to 5 rating on group size and ratio (for which the higher share would apply if the ratio were based on adults in the room rather than staff). Considering just the staff education criteria of the staff education and training element, the data suggest that about half of the programs would reach Tier 4 or 5, for which the proposed standard requires an AA degree or higher. At the same time, using either the ECERS-R or the CLASS to measure the teaching and learning component, at best about one in four programs would reach Tier 4 or 5 given the cut points we assumed for those scales.

The data from the GTQ initiative in San Francisco County further illustrate the quality variation across providers of various types. First, those data show that more center-based programs would reach at least Tier 4 based on the ERS than programs in the statewide data, suggesting some selectivity into GTQ as well as the benefits from targeted QI efforts. Notably, GTQ targets publicly funded programs, such as California State Preschool, other Title 5 programs, and Head Start, programs that have higher quality standards and tend to score higher on some quality dimensions than center-based programs on average (Karoly et al., 2008). At the same time, the GTQ data provide some indication that rooms serving infants and toddlers would not score as well as those serving preschool-age children. However, the bigger quality differential is with FCCHs: Home-based programs score considerably lower on the applicable ERS than center-based programs do.

The bottom line, accounting for the three quality elements included in our statewide data and the proposed cut points, is that fewer than 10 percent of center-based

preschool programs that might participate in a QRIS would be expected to initially score in Tier 5. That share likely would be higher among those programs that would choose to join the system, and it would be expected to rise over time as QI efforts boost quality. Most importantly, of the three quality elements we examined, most programs would find that their scores on the teaching and learning element would hold them to a lower rating tier than the other quality elements would.

The small share of programs that would meet the standards set for Tier 5 is consistent with the advisory committee's goal of making that rating level aspirational. And it is consistent with many other QRISs that have likewise set high standards for the top tier in their rating system (Tout et al., 2010).



In considering these findings, it is important to keep in mind a few limitations of our analysis. First, this analysis is not meant to be a definitive simulation of the ratings that would result from implementing the QRIS proposed by the CAEL QIS Advisory Committee. The available data cover only three of the five quality elements in the proposed block rating system, and one was incomplete; we could not replicate all of the concepts in the staff education and training element. If the unexamined family involvement and program leadership quality elements are ones in which programs score in the higher tiers, they would still be constrained by the lower rating on the teaching and learning element given the QRIS's block structure. Conversely, it is possible that more programs would be rated in lower tiers overall if those two elements are ones in which programs typically rate below their tier rating on the teaching and learning element. Thus, our estimates can be viewed as an upper bound on how programs will likely be distributed across the rating tiers when all five quality elements are considered.

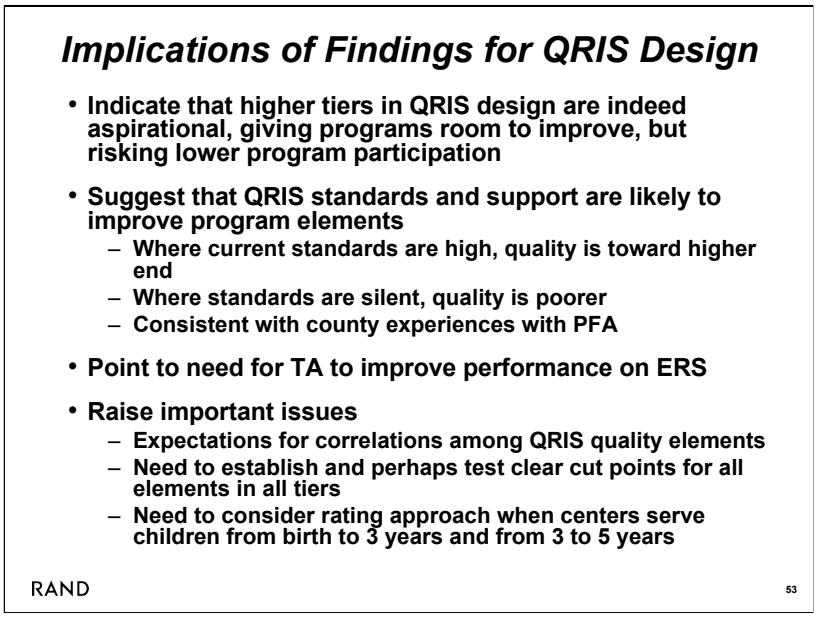
Second, much of our analysis has relied on the statewide data, which include only center-based programs serving preschool-age children. According to the results for the ERS in the GTQ programs, we might expect centers serving infants and toddlers to achieve somewhat lower ratings, while FCCHs would also be concentrated in lower tiers.

Third, our analysis does not account for possible variation in quality across classrooms within the same center, particularly in the statewide data, in which each center is represented by one classroom. In a recent analysis of data from Colorado's QRIS, we have found that the variation within centers in key quality elements is sufficient to affect the rating a program would receive, depending on which classroom or classrooms are assessed (Karoly, Zellman, and Perlman, 2011).

Fourth, the contrast in the results for the statewide data, which are representative of all programs, with those from the GTQ initiative, in which programs are required to

participate if they receive public financial support, demonstrates that the distribution of program ratings is likely to vary in a system that includes a selective set of providers, whether as part of a mandate or in a voluntary system.

Finally, our analysis does not address the validity of the QRIS design in terms of whether the rating levels defined by the system capture meaningful differences in program quality. Further validation research would be required to determine whether the quality elements in the rating system, the specific measures employed, and the associated cut points for each element that determine the rating tiers are predictive of child developmental outcomes (Zellman and Karoly, 2012).



Our findings provide some important support for the value of a QRIS and point to some possible implications for a QRIS design. A key feature of QRISs is that standards for quality are articulated and operationalized in the rating protocol, which identifies the key elements of quality and indicates what it takes to attain each tier rating. The analyses presented in this briefing demonstrate clearly that the proposed QRIS design for California defines a set of rating tiers in which the highest two tiers, especially the highest tier, are aspirational. The finding that few programs would be expected to reach the highest tiers, based on current quality levels, means that there is room for programs to improve over time, as intended. At the same time, if only a small share of programs reached the top tiers, at least initially, some providers may be less inclined to participate in a voluntary system.

Our finding that programs score better on those QRIS elements that are already addressed in standards, such as ratio and group size, is encouraging because it suggests that, when standards are put in place for more quality elements through a QRIS, programs are likely to improve their quality on those elements as well. This has certainly been the experience of counties that have implemented PFA (American Institutes for Research, 2007). For this reason, it is reasonable to expect that, over time, more programs would reach higher rating tiers.

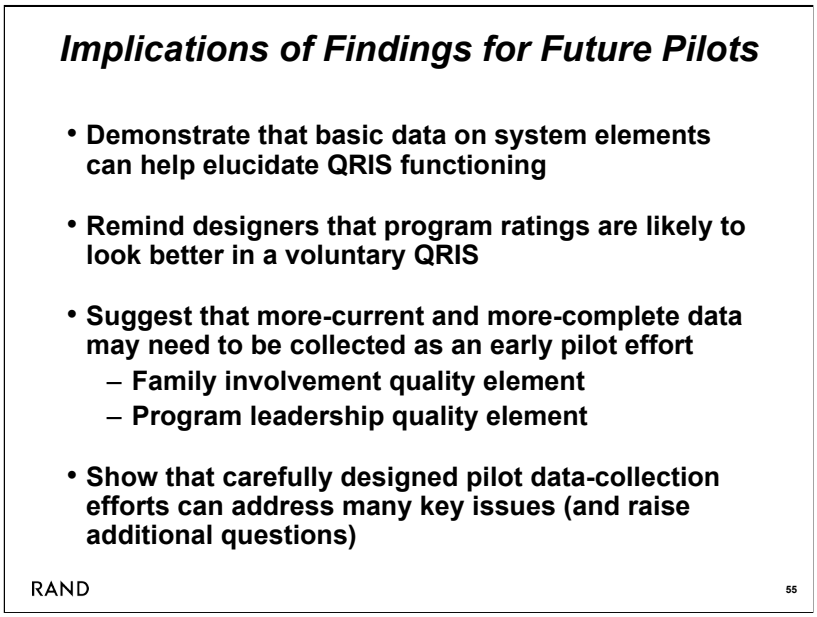
Our data also identify, among the three elements we examined, areas in which programs are likely to need the most assistance in order to raise their overall rating—namely, in improving their performance on the relevant ERS (i.e., ITERS-R, ECERS-R, or FCCERS) or the CLASS (if that comes to be included in the standards). This information may help system planners to design TA approaches that focus on those elements or to reconsider cut points for those elements.

This work may also be helpful in raising issues that have not yet been addressed in the QRIS design process. For example, expectations that quality elements will be

highly correlated may not always hold; we found, for example, that having staff with levels of education that qualify for the highest tiers does not necessarily ensure a program will score at that same level on the relevant ERS or on the CLASS. At the same time, low levels of correlation across rating elements can be an advantage in a QRIS design to the extent that each element captures a distinct dimension of quality, rather than overlapping or redundant quality dimensions (Zellman and Perlman, 2008).

Another key issue is the need to fully specify some currently unspecified aspects of the QRIS design such as the cut points that would apply to the CLASS in Tier 4 or 5. Given the differences in scores on CLASS subscales, ways to combine the subscales and use the CLASS merit further attention.

Finally, the data from the GTQ initiative suggest that center ratings may differ depending on whether the focus is on rooms serving infants and toddlers or those serving preschool-age children. Thus, ways to generate an overall center-level rating, given those within-center differences, need to be addressed.



This study, as a first step prior to field-based pilot work, demonstrates that careful analysis of existing data sets can be a cost-effective and very helpful tool for QRIS designers and administrators. Such work can identify the degree to which system elements are related and reveal how closely the implicit assumptions underlying the rating process align with the realities of program functioning.

At the same time, it is important to recognize that existing data sets have limits that affect the generalizability of the types of analyses presented here. In this case, the programs included in the RAND statewide data set did not volunteer for a QI effort. In contrast, in a voluntary system, QRIS participants choose to participate. Consequently, the distribution of ratings from a voluntary QRIS, as proposed for California, is likely to look better. How much better can be determined only through pilot studies or following full QRIS implementation.

Given what has been learned from this analysis, there are clear benefits from filling the knowledge gaps that remain. For example, as noted earlier, we were not able to assess how programs would rate based on the proposed measures of family involvement and program leadership. Thus, an early pilot effort could measure those quality elements so that these important system components could be examined prior to QRIS implementation.

In the end, the analyses reported here represent a simple and cost-effective way for QRIS designers to obtain a first look at the consequences of the choices they have made about what quality elements to include and where to set cut points, information that can guide expectations about the rating system and shape the next round of decisions. This approach could be applied using other data sets that contain information on the quality elements included in the proposed California QRIS design. Such data may cover the state or be specific to certain geographic locations. Information on the quality of home- and center-based care settings that serve infants

and toddlers or home-based settings that serve preschool-age children would be particularly valuable.

The methodology employed in this study could be used by other states that are contemplating a QRIS or refining an existing system. In the latter case, existing data collected during the program rating process would be ideal to use because they would include all of the QRIS elements and include only those programs that had elected to participate in the QRIS.

References

American Institutes for Research, *Evaluation of Preschool for All Implementation in San Mateo and San Francisco Counties: Year 2 Report*, Palo Alto, Calif., October 29, 2007. As of April 15, 2012: http://www.earlylearningsystems.org/files/PFA%20Year%202%20Report_Final_11_09.pdf

Barnett, W. Steven, Dale J. Epstein, Megan E. Carolan, Jen Fitzgerald, Debra J. Ackerman, and Allison H. Friedman, *The State of Preschool 2010: State Preschool Yearbook*, New Brunswick, N.J.: National Institute for Early Education Research, 2010. As of April 15, 2012: http://nieer.org/publications/state-preschool-2010

CAEL QIS Advisory Committee—*See* California Early Learning Quality Improvement System Advisory Committee.

California Child Care Resource and Referral Network, *2011 California Child Care Portfolio,* San Francisco, Calif., 2011. As of April 15, 2012: http://www.rrnetwork.org/rr-research-in-action/2011-california-child-care.html

California Department of Education, *California Preschool Learning Foundations*, Volume 1, Sacramento, Calif., 2008. As of April 15, 2012: http://www.cde.ca.gov/sp/cd/re/documents/preschoollf.pdf

———, *California Infant/Toddler Learning and Development Foundations*, Sacramento, Calif., 2009. As of April 15, 2012: http://www.cde.ca.gov/sp/cd/re/itf09aavcontents.asp

California Early Learning Quality Improvement System Advisory Committee, *Dream Big for Our Youngest Children: Executive Summary*, Sacramento, Calif., 2010a. As of April 15, 2012: http://www.cde.ca.gov/sp/cd/re/documents/fnlrptexecsummary.pdf

———, *Dream Big for Our Youngest Children: Final Report*, Sacramento, Calif., 2010b. As of April 15, 2012: http://www.cde.ca.gov/sp/cd/re/documents/fnlrpt2010.pdf

Early, Diane, Oscar Barbarin, Donna Bryant, Margaret Burchinal, Florence Chang, Richard Clifford, Gisele Crawford, Wanda Weaver, Carollee Howes, Sharon Ritchie, Marcia Kraft-Sayre, Robert Pianta, and W. Steven Barnett, *Pre-Kindergarten in Eleven States: NCEDL's Multi-State Study of Pre-Kindergarten and Study of State-Wide Early Education Programs (SWEEP)*, Chapel Hill, N.C.: Frank Porter Graham Child Development Institute, March 31, 2005.

Gateway to Quality, "Gateway to Quality," website, undated. As of April 15, 2012: http://gatewaytoquality.sfsu.edu/about.html

Gormley, William T. Jr., and Ted Gayer, "Promoting School Readiness in Oklahoma: An Evaluation of Tulsa's Pre-K Program," *Journal of Human Resources*, Vol. 40, No. 3, Summer 2005, pp. 533–558.

Gormley, William T. Jr., Ted Gayer, Deborah Phillips, and Brittany Dawson, "The Effects of Universal Pre-K on Cognitive Development," *Developmental Psychology*, Vol. 41, No. 6, November 2005, pp. 872–884.

Hamre, Bridget K., and Robert C. Pianta, "Can Instructional and Emotional Support in the First-Grade Classroom Make a Difference for Children at Risk of School Failure?" *Child Development*, Vol. 76, No. 5, 2005, pp. 949–967.

Harms, Thelma, Richard M. Clifford, and Debby Cryer, *Early Childhood Environment Rating Scale–Revised Edition*, New York: Teachers College Press, 2005.

Harms, Thelma, Debby Cryer, and Richard M. Clifford, *Infant/Toddler Environment Rating Scale–Revised Edition*, New York: Teachers College Press, 2006.

———, *Family Child Care Environment Rating Scale–Revised Edition*, New York: Teachers College Press, 2007.

Howes, Carollee, Margaret Burchinal, Robert Pianta, Donna Bryant, Diana Early, Richard Clifford, and Oscar Barbarin, "Ready to Learn? Children's Pre-Academic Achievement in Pre-Kindergarten Programs," *Early Childhood Research Quarterly*, Vol. 23, No. 1, First Quarter 2008, pp. 27–50.

Karoly, Lynn A., *Preschool Adequacy and Efficiency in California: Issues, Policy Options, and Recommendations*, MG-889, Santa Monica, Calif.: RAND Corporation, 2009. As of April 15, 2012: http://www.rand.org/pubs/monographs/MG889.html

Karoly, Lynn A., Bonnie Ghosh-Dastidar, Gail L. Zellman, Michal Perlman, and Lynda Fernyhough, *Prepared to Learn: The Nature and Quality of Early Care and Education for Preschool-Age Children in California*, Santa Monica, Calif.: RAND Corporation, TR-539, 2008. As of April 15, 2012: http://www.rand.org/pubs/technical_reports/TR539.html

Karoly, Lynn A., Elaine Reardon, and Michelle Cho, *Early Care and Education in the Golden State: Publicly Funded Programs Serving California's Preschool-Age Children*, Santa Monica, Calif.: RAND Corporation, TR-538, 2007. As of April 15, 2012: http://www.rand.org/pubs/technical_reports/TR538.html

Karoly, Lynn A., Gail L. Zellman, and Michal Perlman, "Understanding Variation in Classroom Quality Within Early Childhood Centers: Evidence from Colorado's Quality Rating and Improvement System," unpublished draft, November 2011.

NAEYC—*See* National Association for the Education of Young Children.

National Association for the Education of Young Children, *NAEYC Early Childhood Program Standards and Accreditation Criteria: The Mark of Quality in Early Childhood Education*, Washington, D.C., 2007.

National Child Care Information and Technical Assistance Center, *A Profile of Current State Efforts to Improve Quality*, Washington, D.C., 2010.

NCCIC—*See* National Child Care Information and Technical Assistance Center.

Phillips, Deborah A., William T. Gormley, and Amy E. Lowenstein, "Inside the Pre-kindergarten Door: Classroom Climate and Instructional Time Allocation in Tulsa's Pre-K Programs," *Early Childhood Research Quarterly*, Vol. 24, No. 3, 2009, pp. 213–228.

Pianta, Robert C., Karen M. La Paro, and Bridget K. Hamre, *Classroom Assessment Scoring System (CLASS) Manual, Pre-K*, Baltimore, Md.: Paul H. Brookes Publishing, 2008.

Tout, Kathryn, Rebecca Starr, Margaret Soli, Shannon Moodie, Gretchen Kirby, and Kimberly Boller, *Compendium of Quality Rating Systems and Evaluations*, Washington, D.C.: Administration for Children and Families, Department of Health and Human Services, April 2010. As of April 15, 2012: http://www.acf.hhs.gov/programs/opre/cc/childcare_quality/compendium_qrs/qrs_compendium_final.pdf

Zellman, Gail L., and Lynn A. Karoly, *Moving to Outcomes: Approaches to Incorporating Child Assessments into State Early Childhood Quality Rating and Improvement Systems*, Santa Monica, Calif.: RAND Corporation, OP-364, 2012. As of April 15, 2012: http://www.rand.org/pubs/occasional_papers/OP364.html

Zellman, Gail L., and Michal Perlman, *Child-Care Quality Rating and Improvement Systems in Five Pioneer States: Implementation Issues and Lessons Learned*, Santa Monica, Calif.: RAND Corporation, MG-795, 2008. As of April 15, 2012: http://www.rand.org/pubs/monographs/MG795.html