



Refining Delaware Stars for Early Success: Workshop for the Office of Early Learning

Gretchen Kirby's Presentation on Influences on What QRIS Rating Levels Mean and the Ratings Programs Receive

Wednesday, February 6, 2019

9:00 a.m.–4:30 p.m.

Workshop transcript

Gretchen Kirby:

I really just want to start by giving you all a round of applause. You all should applaud yourselves because you all shared—you really have had a lot of successes in 10 years of doing Delaware Stars.

At this point, there's forty-some QRIS across the country. Lots are statewide; some are at the district level. The systems are no longer new. Right? But there aren't a lot of clear answers about the next directions, and I think that's the point Delaware is at and a lot of states are at is okay, so we've been around for a while. What do we do next? And the reason that this is so complicated is because there is this constellation of policies and structures that vary across the states.

As you know, all the different systems do things in different ways. We don't know that any two systems are exactly alike. The rating structures are different. The number of rating levels can vary. The types of incentives. The requirements for which programs are eligible to participate or that are mandated to participate. Entry policies. I mean you know the array of policies and structures that go into these things. It's a researcher's nightmare because we can't isolate the effect of any one thing. So when you ask us, "What should we do?" we're like well, you know, it's hard because it is a package. So it is hard to get to those really super evidence-based findings of, "This is what you should do. Here's the magic bullet."

There's a lot of descriptive information. There's a lot of suggestive information about ways to go. And that's what we're going to try to help make sense of today. But as you know, there's limited evidence that the differences in quality rating levels translate into differences in child outcomes. And this is where Delaware is. This is where a lot of states are,

coming out of Race to the Top—Early Learning Challenge validation studies. And we're like gosh, we really believe in these systems. Right? We really know or feel that these are making gains for kids and families and systems, the early learning systems. But how can we move the needle on child outcomes? So we're going to focus a lot on those issues this morning.

Despite this lack of clear answers, quality rating improvement systems really have shined a light on important issues about supporting quality in programs that serve young children and have really helped spark a national conversation about how to improve quality to the benefit of all children. When you think even about the subsidy system from 20 years ago, quality was not really a word that was heard very much. And so I think we have to recognize that that's a huge conversational shift to really think about quality in all contexts.

Okay. So what is QRIS? And I don't mean we're going to get into the details of defining every aspect of it. But we need to recognize that QRIS has multiple goals. There's goals for children. There's goals for families. There's goals for programs. And there's goals potentially for system change or integration of programs across the early care and education system.

As a result, a quality rating improvement system can be a measure of quality in the care and education of young children. It can be a tool for families to select care. It can be a performance or improvement framework for programs. And it can be a system of supports for quality improvement.

Wow. That's kind of exhausting, right? That's really challenging to be all of these things effectively. And I think part of the core struggle of QRIS is how do you do all of these things really well? And that speaks to sort of the challenges and the tradeoffs and the hard decisions that sometimes need to be made to make one of these a priority, potentially, over some others. How do you achieve the goals in concert with each other?

So this morning we're really going to focus on QRIS as a measure of quality, bearing in mind the goal to restructure Delaware Stars to distinguish levels of quality in settings and ensure the validity of the ratings. Right? So specifically, the programs that earn higher ratings really lead to the better outcomes for children.

Okay. What are the challenges to doing this? First of all, we need to sort of recognize that rating levels are signals of quality. What do we mean by that? Well, they're generally not created with principles of scale development in mind. Right? So they're not measures in the sense of a validated test like all kids take. They're not a classroom observational, like the Environment Rating Scale or CLASS that's administered the same way consistently across classrooms. So that's a challenge because you're trying to validate something as a measure that maybe isn't structured as a measure.

It allows different paths to quality because you want to engage all programs. There's the goals for participation. There's the goals for programs. So you want to sort of meet programs where they are and help them improve. That, again, can be a challenge from a measure development or measure of use perspective.

Programs can have the ability to select a rating to which they want to apply. So again, it's sort of the meeting programs where they are can introduce some inconsistencies in how you are sort of measuring program quality.

The other challenge is that there's policies and structures of the system that are intended to achieve other goals that I just talked about for programs, for families, for system integration.

So it's not just that the standards work in isolation. Right? It's all of the things that surround how programs enter and move through the system. So there's policies around participation. There's policies such as, again, which programs are eligible to apply, whether they're mandatory or voluntary, whether there's automatic ratings at which some programs enter at higher levels, the incentives around the ratings that are tied to participation or movement. And then quality improvement policies that are structured to move programs from lower to higher ratings, such as tiered reimbursement that we'll talk about this afternoon.

And then there's other influences, again, some policy, some structural. Entry policies that could be based on program type or funding. So are all programs required to sort of enter at, say, Starting with Stars, or Level 1 in some states, or are programs that receive child care and development funds required to be at a certain level, say, 3 or higher, et cetera? The rating structure is it points, hybrid, building block, and the validity period of ratings. So all of these things sort of play into how QRIS—the challenges to how QRIS can be thought of as a measure of quality.

Switching from sort of the challenges of QRIS as a measure, let's think about what does that introduce in terms of challenges for the validity of quality rating levels, for distinguishing outcomes for kids between lower- and higher-rated programs. Okay. We all know that there's a large number of indicators, typically, in QRIS, and those are really useful for program performance, to guide improvement. As a performance measurement framework, all of those pieces are important. They're research based. There's a reason that they're all there. Because we want programs to focus on different elements of the care setting and the care that they provide to young children.

The flip side is that you have a lot of indicators, but a parsimonious set, meaning a very slim set of indicators that's closely related to children's

experiences in the classroom is going to give you sort of the best thing for your buck in affecting child outcomes. So focusing in on the set of indicators, as I said, that can really affect child outcomes is sort of something that could produce a better measure and lead to validity, but can be juxtaposed against other goals that you have within the system.

The other challenge to the validity that we saw in Delaware and across other states is that there's limited variation of quality across the rating levels, so it's hard to distinguish the quality between each rating level. And this, in part, can be due to program responses to these system characteristics and policies that I just talked about. Right?

So entry policies. Programs of higher quality may enter at low rating levels. So the problem for the validity of the system is that if you have this mishmash, sort of, of quality within a level, you're just not going to be able to find those distinctions in kids' outcomes. Right? So it's not always just about the standards. You have to sort of think about all the other pieces that are going into it that are going to help you really hone in on the differences in quality and how that can affect kids.

So, as you said, the rating structure is another example of that, block structures, in particular. A lot of programs could meet most of the elements and just be a little bit shy of one. So they may be in a rating level that is lower than really what their true measured quality could actually be.

And then again, there's the incentives and validity periods. They're shorter. They can be shorter or longer validity periods. But I think Martha or Heidi just mentioned sort of that timing issue. Programs can change quality and, unfortunately, could change relatively quickly.

And then lastly, the average levels of quality just may not be high enough to affect children's outcomes. Like are we really getting programs at the highest level to where they could be or should be to get to a point that you can affect kids' outcomes?

We can see that centers participate in Delaware Stars at much higher rates than family child care providers. We do see a slight increase in center participation from 2014 to 2017, but the general picture remains the same. Right? Relatively higher center participation and low family child care participation.

I should note that centers don't include the school-aged programs and family child care includes both the large and the small.

So just to give a sense of the full picture of participation, again, this includes all programs and, again, sort of emphasizes that licensed centers are the largest share, but this one also includes school aged, Head Start, and the ECAP programs. So to clarify, Head Start and ECAP are mandated

participants in Head Start, and they have pretty close to 100 percent participation, but they're relatively small numbers, so they still comprise a relatively small portion of the pie.

So looking at the distribution of ratings over time—this is all programs including center-based and family child care—you can see the distribution of rating levels really has shifted from the lower to the highest ratings, which is something that we tend to see in the more mature programs.

Think about the standards but also what we've talked about in terms of the context of the policies and how programs enter and move through the system, the other influences that are occurring. And can you think about sort of one opportunity that you see in making this revision that's going to sort of address what you're trying to get at in terms of consistency in quality and distinctions in quality and one challenge that could be, again, sort of related to the standards or related to some of the policies or other structural aspects.

This work was funded by the U.S. Department of Education's Institute of Education Sciences (IES) under contract ED-IES-17-C-0006, with REL Mid-Atlantic, administered by Mathematica Policy Research. The content does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.