



Measuring the quality of teacher–child interactions at scale: Comparing research-based and state observation approaches



Virginia E. Vitiello*, Daphna Bassok, Bridget K. Hamre, Daniel Player, Amanda P. Williford

University of Virginia, United States

ARTICLE INFO

Article history:

Received 14 April 2017

Received in revised form 27 February 2018

Accepted 5 March 2018

Keywords:

Classroom observations

Qris

Preschool

ABSTRACT

Use of observational measures to monitor preschool quality is growing rapidly. Although a large body of research has examined the validity of classroom observation tools within the context of researcher-conducted studies, little research to date has examined the extent to which the observations conducted as a part of state accountability efforts correspond to observations collected by research teams. This paper examines the degree of agreement between local and research rater teams using an observational measure of preschool classroom quality. It also explores the extent to which ratings predicted gains in children's literacy, math, and self-regulation skills. Local ratings were conducted as a part of Louisiana's quality rating and improvement system. Both rating teams observed 85 classrooms offering publicly funded preschool programs using the Pre-K CLASS, and 820 children from these classrooms (average age = 52.6 months, SD = 3.6 months) were directly assessed in the fall and spring. Results indicated correlations between local and research teams' scores on corresponding domains, ranging from $r = .21$ to $.43$. Both teams' scores were significantly but modestly related to children's learning gains, although patterns of association differed. Results are discussed in the context of policies that require observational measures at scale.

© 2018 Elsevier Inc. All rights reserved.

Early childhood education (ECE) programs can yield short and long-term benefits for children (Phillips et al., 2017). However, many children in the United States attend ECE programs that do not offer high quality environments (Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Dowsett, Huston, Imes, & Gennerian, 2008). Lower-quality programs are less likely to benefit children in terms of developing school readiness skills than are higher-quality programs (Karoly, 2014; Sabol & Pianta, 2014, 2015). This has led to substantial public investments in improving ECE quality. Early childhood accountability systems have become one increasingly prominent policy lever. Spurred by Federal funding from the Race to the Top Early Learning Challenge, today nearly all states have developed and are expanding Quality Rating and Improvement Systems (QRIS; *The Build Initiative and Child Trends*, 2016). QRIS are accountability systems, typically administered at the state level that define quality benchmarks for ECE programs and seek to improve quality both through supports and incentives for programs

and by providing parents with information about ECE program quality, to help them make informed choices.

It is not yet clear whether public investments in QRIS are leading to meaningful system-wide improvements in ECE program quality. One concern is that the rapid design and rollout of states' QRIS systems has outpaced the research base around accurately measuring quality. In order to lead to quality improvements – and, ultimately, better child outcomes – QRIS must begin by accurately measuring the features of program quality that affect child learning (Cannon, Zellman, Karoly & Schwartz, 2017).

However, despite decades of research on efforts to measure quality in ECE settings, many questions remain about how to do this accurately at scale (Burchinal, 2017). For instance, most states include classroom observations as a component of their QRIS (*The Build Initiative and Child Trends*, 2016), in part because a large body of evidence demonstrates positive, though modest, associations between these classroom observations and children's learning. However, there is relatively little evidence about the use of classroom observations at scale for policy applications like QRIS (Goffin & Barnett, 2015). It is not yet clear whether measures of classroom quality collected as a part of large-scale policy initiatives capture

* Corresponding author at: University of Virginia, Center for Advanced Study of Teaching and Learning, P.O. Box 800784, Charlottesville, VA 22904, United States.

E-mail address: vev9m@virginia.edu (V.E. Vitiello).

child outcomes as well as do measures collected by researcher-based teams, especially when quality ratings are tied to stakes.

Given the United States' growing investments in classroom observations, which can be costly and time-consuming to collect, it is important to address this gap. Using data from Louisiana, we compare classroom observations collected by local raters to observations conducted by independent data collectors using a standard research protocol. Both teams observed classrooms using the Classroom Assessment Scoring System (CLASS), a widely used measure of teacher–child interactions (Pianta, La Paro, & Hamre, 2008). Their approaches differed, however, in that the research team observed each classroom more frequently, and was also more explicitly focused on strategies to ensure rater reliability. The goal of this study is to assess to what extent the observations conducted according to these different approaches lead to similar conclusions about program quality. The results of this exploratory analysis raise considerations for policy makers determining how to include classroom observations into accountability systems.

1. QRIS & efforts to measure and improve quality at scale

By articulating a clear definition of quality, measuring programs' performance relative to that definition, and providing programs with incentives and supports, QRIS aim to create a culture of improvement (Goffin & Barnett, 2015; Zellman, Perlman, Le, & Setodji, 2008). Because QRIS are relatively new and because in many states they are not implemented at scale, there has not yet been research on the effects of these accountability systems on children's learning.

Several studies do provide encouraging evidence that QRIS can foster program-level changes. For instance, one small randomized control trial in Washington State demonstrated that programs participating in a QRIS with coaching supports demonstrated increases in quality as measured using a widely used observational tool (Boller et al., 2015). A recent study of North Carolina's QRIS system found that quasi-random assignment to a lower quality rating led programs to make notable improvements on a multi-faceted measure of classroom quality (Bassok, Dee, & Latham, 2017).

These studies suggest that at least in some contexts, ECE programs are responsive to the incentives and supports embedded in QRIS. However, for QRISs to foster meaningful change, it must be the case that they define and measure quality in a way that is closely aligned with children's development.

Most QRIS rate programs based on a complex set of factors including structural features (e.g. class size, ratios, teacher credentials), classroom observations (e.g. using Classroom Assessment Scoring System (CLASS; Pianta et al., 2008) or the Early Childhood Environment Rating Scales (ECERS; Harms, Clifford, & Cryer, 1998)), and a host of other measures (e.g. family engagement, administration and business practices, measures of curriculum and assessment use, etc.). States typically use some formula to combine these disparate metrics into a single quality rating, which is typically broken into 3–5 quality levels. Programs scoring above high-level thresholds are publicly recognized as high quality programs in ways that are intended to drive greater enrollment. They often also receive fiscal rewards. Programs scoring at very low levels may receive additional professional development and/or have more punitive sanctions such as a reduction in subsidies.

Ensuring alignment between quality ratings and child outcomes is so central to the QRIS theory of change that, to date, the vast majority of QRIS research has focused on rating validation (Goffin & Barnett, 2015). Existing research suggests that many of the individual metrics included in QRIS are weak predictors of children's learning in ECE settings (Early, Maxwell, Ponder, & Pan, 2017; Mashburn et al., 2008) and that they are not systematically related

to child outcomes when grouped together and used to create program ratings (Sabol, Hong, Pianta, & Burchinal, 2013). Further, a growing body of QRIS validation studies has generally found no or inconsistent associations between QRIS ratings and children's outcomes (Cannon et al., 2017; Karoly, 2014).

Across 15 recent reports, four found no differences by QRIS rating (Sirinides, 2010; Tout et al., 2010; Tout, Starr, Albertson-Junkans, Soli, & Quinn, 2011; Zellman et al., 2008), and the rest found small, non-linear associations, that are typically significant for just one skill domain (e.g. Elicker, Langhill, Ruprecht, Lewsader, & Anderson, 2011; Sirinides, Fantuzzo, LeBoeuf, Barghaus, & Fink, 2015; Soderberg, Joseph, Stull, & Hassairi, 2016; Thornburg, Mayfield, Hawks, & Fuger, 2009; Tout et al., 2016).

This lack of predictive validity is a serious threat to the utility of QRIS, and has led to calls for new ways of measuring quality in ECE settings, especially at scale (Burchinal, 2017; Cannon et al., 2017; Karoly, 2014). These calls have focused on the need to simplify quality ratings by focusing on fewer measures that have consistent, demonstrable links with children's learning (Sabol et al., 2013; Sabol & Pianta, 2015).

2. Classroom observations as a tool for quality measurement

One potentially promising quality measure for large-scale accountability systems is the CLASS, a widely used observational measure of teacher–child interactions that assesses effective interactions across ten dimensions divided into three broad domains: Emotional Support, Classroom Organization, and Instructional Support (Pianta et al., 2008). Currently 45% of QRIS systems use the CLASS (The Build Initiative and Child Trends, 2014) and it is also included in Head Start's monitoring system, the Designation Renewal System (DRS; Administration of Children and Families (ACF), 2011). A substantial research base shows a positive relationship between CLASS scores and gains in child outcomes, although these relationships are typically small.

For instance, research using the CLASS indicates that when teachers offer warm, supportive, and responsive interactions, children develop stronger social and emotional skills (e.g., Johnson, Seidenfeld, Izard, & Kobak, 2013). Children in classrooms with strong behavior management and classroom organization demonstrate stronger growth in self-regulation skills (Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009). Further, teachers' daily provision of cognitively stimulating instruction and conversation appears to be a critical ingredient in fostering academic learning (e.g., Howes et al., 2008). Most compellingly, a recent experiment that randomized children to classrooms within schools showed that young children make greater gains in language, math, and executive functioning skills in classrooms where teachers were more highly rated on the CLASS (Araujo, Carneiro, Cruz-Aguayo, & Schady, 2016).

Although the positive relationship between CLASS and various child outcomes has been documented widely – and has motivated many states to include the measure in their QRIS – many questions about the role of CLASS within large-scale accountability remain. First, the associations between CLASS scores and child outcomes tend to be modest. Araujo et al. (2016) found that a standard deviation increase on the CLASS was associated with .07–.11 standard deviation increases in child outcomes; Keys et al. (2013) used meta-analytic techniques across multiple studies using various observation measures of quality, including the CLASS, and found an average standardized main effect of .05 on child outcomes. This is not a problem unique to CLASS, however; as a field, we do not yet have measures of quality that are systematically related to moderate or large increases in child outcomes. Still, the relatively

Download English Version:

<https://daneshyari.com/en/article/6840574>

Download Persian Version:

<https://daneshyari.com/article/6840574>

[Daneshyari.com](https://daneshyari.com)