Accountability Comes to Preschool: Florida’s Approach to Evaluating Pre-Kindergarten Programs Based on Their Graduates’ Kindergarten Assessments

Luke C. Miller, Daphna Bassok, Amanda J. Johnson & Eva Galdo
Introduction

High-quality preschool programs help students start school “ready to learn” and can yield a host of important long-term benefits. The quality of child care and preschool programs nationwide, however, is highly variable. Early childhood accountability systems, often known as Quality Rating and Improvement Systems (QRIS) are one increasingly popular approach for improving quality in early childhood settings. There are 39 states with a state-wide QRIS, and nearly all others are in the planning or piloting phases. The goal of these early childhood accountability systems is to accurately measure program quality, incentivize improvements, and provide parents with information about preschool quality so they can make informed choices.

An important open question about early childhood accountability systems is what exactly states should measure and incentivize. Many state systems focus on a combination of structural characteristics (e.g., child:adult ratios, teacher education levels) and classroom observations, which aim to capture the quality of interactions between children and their caregivers. This approach to program accountability, which is focused on inputs and processes rather than on children’s outcomes, stands in contrast to the K-12 sector where teacher- and school-level accountability systems are often tied to outcomes, and most commonly student learning gains on standardized tests.

Accurately assessing children’s learning gains in early childhood settings is notoriously costly and problematic. Indeed, a number of reports have outlined the many unique challenges associated with assessing very young children and have cautioned strongly against the use of child assessments as part of summative or high-stakes evaluations of programs serving young children.\(^1\)

On the other hand, a mounting body of evidence raises concerns that the types of metrics usually used to evaluate early childhood classroom quality in accountability systems are, at best, weakly associated with children’s learning.\(^2\) As public investments in early childhood grow, there is a heightened need to measure program effectiveness, to target professional development supports to programs in need, and to ensure children have access to effective early childhood environments. Some worry that defining “high-quality” without explicitly accounting for measures of children’s development will provide weak and ineffective tools.\(^3\)

This policy brief describes one state’s experience using child assessment data to evaluate the quality of early childhood providers. In 2005, Florida introduced its Voluntary Prekindergarten (VPK) program, a free, universal preschool initiative. VPK currently serves about 75 percent of the state’s 4-year-olds, and is a national leader with respect to preschool access. Since its inception, Florida has evaluated the quality of VPK programs based on their participants’ score on an assessment administered at the beginning of kindergarten. This memo describes Florida’s unique approach to program-level accountability and also highlights some potential unintended consequences of Florida’s early childhood accountability system. In particular, we focus on the lack of a pre-test and the related possibility of mislabeling programs “low performing.”

Acknowledgements: This research was funded by the Annie E. Casey Foundation, the Foundation for Child Development, and the Smith Richardson Foundation. It was also supported by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305B090002 to the University of Virginia. We thank them for their support and acknowledge that the findings and conclusions presented in this report are those of the authors alone, and do not necessarily reflect the opinions of these foundations, the Institute, or the U.S. Department of Education.

Measuring preschool quality in Florida’s VPK Program

Like all states with public preschool programs, Florida ensures their programs meet a basic level of quality through professional development requirements for teachers, maximum class size restrictions, the mandated use of developmentally appropriate curriculum, and licensing and accreditation requirements. Florida is unique, however, in that it also evaluates programs based on their students’ test scores. State law requires that all incoming kindergarteners in the public schools
be assessed within the first 30 days of kindergarten.\textsuperscript{4} In addition, all children who participate in VPK but attend a private school for kindergarten must also be assessed, a process that the local school district coordinates. After kindergarteners are assessed, their scores are linked back to the VPK program they attended. Programs are then rated based on the aggregated scores of their graduates.

**What does the Florida Kindergarten Readiness Screener measure?**
The Florida Kindergarten Readiness Screener (FLKRS) collects information on children’s development and aims to measure the readiness of each child for kindergarten based on the Florida Early Learning and Development Standards for 4-year-olds. As shown in Table 1, the specific set of assessments administered as part of the FLKRS has changed over the years but have always consisted of observational and cognitive assessments. Initially, three screening measures were used: the Early Childhood Observation System (ECHOS), which measures a child’s development in seven areas: language and literacy, mathematics, social and personal skills, science, social studies, physical development and fitness, and creative arts; the Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Letter Naming Fluency; and, DIBELS: Initial Sound Fluency. In the fall of 2009, the DIBELS exams were replaced by the Florida Assessments for Instruction in Reading (FAIR) which consists of letter-naming and phonemic awareness tasks that are intended to measure emergent reading skills. Finally, in 2014-15 the Work Sampling System replaced ECHOS and is another observational tool designed to capture a child’s development across many of the same domains as ECHOS.

**How are “low performing” programs identified?**
After children are assessed using the FLKRS in the fall of their kindergarten year, their scores are linked back to the VPK program that they attended. FLKRS scores are then aggregated in order to assign the VPK programs a kindergarten readiness rate based solely on their graduates’ beginning-of-kindergarten FLKRS scores.

Although Florida has made several adjustments to both the assessments administered and to how VPK programs’ readiness rates are calculated (Table 1), the rates have always reflected the percent of children deemed as “kindergarten ready” on the various assessments administered. The kindergarten readi-

<table>
<thead>
<tr>
<th>Year</th>
<th>Assessments Used</th>
<th>Readiness Rate Calculation</th>
<th>Scale</th>
<th>Low-Performing Cutoff Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005-06</td>
<td>1) Early Childhood Observation System (ECHOS)</td>
<td>Sum of the percent of children ready on each of the three screening measures</td>
<td>0 – 300</td>
<td>211</td>
</tr>
<tr>
<td>2006-07</td>
<td>2) Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Letter Naming Fluency</td>
<td></td>
<td>0 – 300</td>
<td>211</td>
</tr>
<tr>
<td>2007-08</td>
<td>3) DIBELS: Initial Sound Fluency</td>
<td></td>
<td>0 – 300</td>
<td>214</td>
</tr>
<tr>
<td>2008-09</td>
<td>1) ECHOS</td>
<td>Sum of the percent ready on each of the two screening measures</td>
<td>0 – 200</td>
<td>138</td>
</tr>
<tr>
<td>2009-10</td>
<td>2) Florida Assessments for Instruction in Reading (FAIR)</td>
<td></td>
<td>0 – 100</td>
<td>140</td>
</tr>
<tr>
<td>2010-11</td>
<td>1) ECHOS</td>
<td></td>
<td>0 – 100</td>
<td>70</td>
</tr>
<tr>
<td>2011-12</td>
<td>2) Florida Assessments for Instruction in Reading (FAIR)</td>
<td></td>
<td>0 – 100</td>
<td>70</td>
</tr>
<tr>
<td>2012-13</td>
<td>1) Working Sampling System (WSS)</td>
<td></td>
<td>0 – 100</td>
<td>70</td>
</tr>
<tr>
<td>2013-14</td>
<td>2) Florida Assessments for Instruction in Reading aligned to Florida Standards (FAIR-FS)</td>
<td></td>
<td>0 – 100</td>
<td>70</td>
</tr>
<tr>
<td>2014-15</td>
<td>1) Working Sampling System (WSS)</td>
<td></td>
<td>0 – 100</td>
<td>70</td>
</tr>
</tbody>
</table>

\textsuperscript{4} No VPK Provider Kindergarten Readiness Rates are available for 2013-14 as the State did not administer the FAIR.
ness rate in VPK’s first years, for instance, was calculated by summing across the three assessments the percentage of a program’s graduates that were deemed kindergarten ready. Children who scored “consistently demonstrating” or “emerging/progressing” on the ECHOS or who scored “above average” or “low risk” on the DIBELS screening measures were classified as ready with respect to that particular assessment. A VPK program with 55, 65, and 75 percent of its graduates deemed ready on the three assessments, respectively, would be assigned a kindergarten readiness rate of 205 (55+65+75). Starting in 2008-09, when the FAIR replaced the two DIBELS assessments, the readiness rates were calculated as a sum across two assessments. What proved to be a more significant change occurred in 2010-11 when the readiness rate was calculated as the percent of children who qualified as ready on both of the screening measures.

Not all VPK programs receive a readiness rating, however. A VPK program must have at least four children with complete FLKRS scores who completed at least 70 percent of the program’s total hours. Programs without sufficient data are therefore not subject to Florida’s accountability system. As shown in Figure 1, roughly 20 percent of programs were excluded in the first year of the VPK program, and since then roughly 10 percent of VPK programs have not received a readiness rating.

Once each program receives a kindergarten readiness rating, programs with the lowest readiness rates are designated as low performing. The cutoff score for this designation is determined by the State Board of Education. The cutoff score, too, has fluctuated over the years due in part to the changing scale of the readiness rating and in part to the changes in how the ratings are calculated (Table 1). During the program’s first five years, the Board was required to set a cutoff score so that no more than 15 percent of VPK programs (with a readiness rating) were rated low performing. The cutoff score generally implied that low-performing programs were those programs where fewer than 70 percent of graduates were deemed ready on the average FLKRS assessment (e.g., 211/3 = 70.3 in 2006-07, 138/2 = 69 in 2008-09). Beginning in 2010-11, the cutoff score for low-performing providers was fixed at 70 percent of graduated deemed ready on both FLKRS assessments.

This 2010-11 change had a significant impact on the percent of VPK programs labeled low performing (Figure 1). Whereas the percent of programs labeled low performing remained just under 15 percent (the maximum allowed) over the program’s first five years, it spiked to nearly 35 percent in 2010-11. Two years later it had dropped to just over 20 percent.

![Figure 1. Percent of programs labeled low performing, 2005-06 to 2012-13](image)
Does labeling program based on their students’ performance make sense?

According to VPK program documentation, the kindergarten readiness rate “measures how well a VPK provider prepares four-year-olds to be ready for kindergarten based upon the Florida Early Learning and Developmental Standards for Four-Year Olds.”

This claim, that the readiness rate is a meaningful measure of a preschool program’s effectiveness, is untrue and likely misleading.

Children enter preschool programs with widely varying abilities, interests, and backgrounds. Without accounting for differences in children’s skills when they enter a VPK program, it is impossible to attribute variation in school readiness at kindergarten to the quality of preschool programs. Imagine two preschool programs. The first, Program 1, serves very low-income children in a neighborhood characterized by high levels of poverty and very limited social services. The second program, Program 2, operates in an affluent community and serves children who have already had many enriching early learning opportunities. If we compare the kindergarten readiness assessments of children who attended each of these programs, we would not be surprised to find that the children at Program 1 scored lower than their peers at Program 2. It would be unwise, however, to assume these gaps are caused by differences in preschool program quality. In fact, when we observe that the kindergarteners who attended Program 1 performed less well than those in Program 2, we have no way to know whether Program 1 is less, more, or equally effective.

In other words, the VPK readiness rate does not measure the value-added of attending a particular VPK provider. Rather than isolating the unique contribution of each VPK program to its graduates’ kindergarten readiness, the readiness rate is capturing—at least in part—differences across students that existed prior to VPK participation.

This problematic feature of the readiness rating leads to a set of potentially unintended consequences. First, providers serving the most at-risk children will be more likely to be labeled low performing irrespective of the program’s true effectiveness. Second, the design of the accountability system may leave providers less inclined to operate in areas that serve the children with the greatest needs, thus limiting access in communities that serve at-risk children. Below we briefly explore both of these concerns.

Preschool programs serving the most low-income and non-white children are far more likely to be rated “low performing”

A major critique of Florida’s accountability system is that it may disproportionately label programs serving low-income and non-white children as low performing. Our analysis provides suggestive evidence consistent with this claim (Figure 2). There are substantial disparities in the likelihood a program will be classified as low performing by the percentage of a VPK program’s graduates that are Black or Hispanic or the percentage eligible for free or reduced-price lunch.

In 2005, roughly 6 percent of programs in the lowest quartile with respect to percentage of Black or Hispanic students were classified as low performing. In the top quartile, nearly 27 percent of programs were labeled low performing. By 2008 this gap had grown even wider. The likelihood of being classified as low performing was seven times higher in programs serving the most Black and Hispanic children.

Disaggregating based on socio-economic status highlights even larger disparities. In 2008, in programs serving the fewest children eligible for free or reduced price lunch, just over 2 percent were labeled low performing. In contrast, nearly 33 percent of programs serving the highest proportion of free or reduced-price lunch eligible children were rated as low-performing.

Our results indicate that the likelihood of being labeled low performing varies drastically depending on programs’ socio-economic and racial composition. It is also the case that the communities surrounding elementary schools that serve the most Black, Hispanic, and low-income children have a far higher rate of VPK programs classified as low performing. For each public elementary school in Florida, we calculated the percentage of VPK programs within a 10-mile radius that were classified as low performing. We then examined whether these percentages differed depending on the racial composition or socio-economic status of...
the children served in the public school. The results are clear. The higher the percentage Black or Hispanic children served in an elementary school the higher the percentage of VPK programs classified as low-performing in the surrounding community (Figure 3). The same patterns hold for free or reduced-price lunch eligibility. It is worth noting that these large disparities could certainly reflect some true differences in the quality of programs serving the most low-income, Black, and Hispanic children. Existing studies have documented meaningful quality differences based on the racial and socio-economic composition of a program or community. The problem is that as currently constructed there is no way to determine to what extent the large disparities in readiness rates are capturing these real differences in quality and to what extent they are only capturing differences in the type of children served.

**Lessons from Florida about early childhood accountability**

Early childhood accountability systems strive to improve children’s learning opportunities by (1) accurately measuring program quality in preschool settings, (2) incentivizing provider improvements, and (3) providing parents with information about preschool quality so they can make informed choices.

Florida’s accountability system has the second and third components at least partially covered. There are incentives for providers to exit the low performing designation. Providers designated low performing are placed on probation, they must submit an improvement plan, and must adopt a curriculum approved by the Department of Education. Those who fall below the minimum readiness rate for multiple years lose their eligibility to provide VPK services.⁸

There are also efforts to make readiness rates and the low performing designation readily available to parents. Kindergarten readiness rates and low performing statuses are publicly available on the Florida Department of Education Office of Early Learning website. Many of the Early Learning Coalitions (ELCs), which are regional organizations responsible for delivering local early learning services, provide searchable databases listing the ratings of VPK providers. In addition, media outlets publicize these programs often providing lists of the “best” and “worst” programs based on the ratings.

The system fails, however, when it comes to accurately measuring program quality. It evaluates programs in a way that likely penalizes programs serving the children most in need of services and may also create powerful disincentives for providers to operate in communities with large populations of
non-white and low-income children.

States are increasingly assessing the school readiness of incoming kindergarteners, and it is not surprising that policy makers interested in improving the quality of early childhood are intrigued by the possibility of using these kindergarten assessments as potential tools for measuring and improving preschool quality.

At the same time, the use of kindergarten assessments to evaluate early childhood programs is uncommon and contentious for a host of reasons that range from poor psychometric properties, developmental appropriateness, cost, and feasibility. These are important concerns worthy of further investigation. There may not be reliable, valid, developmentally appropriate, and feasible ways to assess the effectiveness of early childhood programs using child assessment data. Finding workable solutions will likely require compromising across competing demands. However, Florida’s accountability system from 2005 to 2012 compromised too much with respect to the accuracy and fairness of its quality measure.  

Figure 3. Percentage of VPK programs labeled “low performing” by the racial and socio-economic composition of the surrounding community, 2005 to 2010

<table>
<thead>
<tr>
<th>Year</th>
<th>Lowest Quartile</th>
<th>2nd Quartile</th>
<th>3rd Quartile</th>
<th>Highest Quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


5 https://vpk.fldoe.org/infopages/faq.aspx

6 We calculate each VPK program’s racial composition and its free and reduced price lunch eligibility based on all program participants who are observed enrolled in kindergarten in any public school in the state the following year. Our estimates of each programs’ demographic composition may be biased by our exclusion of children who attend private school or leave the state in the year prior to kindergarten.

7 We do not have student-level demographic data on VPK program participation after 2008.

8 Beginning in 2010, a provider about to lose its eligibility may apply for and potentially receive a “good cause” one-year renewable exemption from the State Board of Education. Evidence of recent
learning gains is mentioned in the statute as an example of a good cause. During this time, the provider is required to maintain state health and safety requirements and serve at least two times the state-wide percentage of children with disabilities or children learning English (Statutes of Florida, ch. 1002 § 67, 2010).

Starting in the 2012-13 school year, VPK programs were required to conduct fall and spring child assessments using the Florida VPK assessment. However, these assessments are meant to be used only as formative assessments and there is no provision in law to use the results of this assessment in the VPK Provider Kindergarten Readiness Rate calculation. Kindergarten readiness rates were not calculated since 2013-14, and it is unclear how Florida intends to calculate the readiness rate moving forward.