

The Quality Conundrum

What is “high-quality” preschool? Nobody wants little kids placed in shoddy programs. But a raging dispute surrounds the definition of quality in this field, and the dominant versions are ill suited to the modern kindergarten-readiness era. Those who would “leave no toddler behind” have enormous difficulty spelling out the preschool equivalent of what, in the K–12 sector, we typically term “proficiency.” And they seem loath to devise and deploy suitable measures of progress toward that result.

Indeed, I sometimes feel as if pre-K advocates inhabit a different universe from the K–12 policy world where I mainly dwell. Two fixed stars are missing altogether: States have no constitutional duty to provide their citizens with preschool education (or didn’t until advocates persuaded Georgia and Florida to amend their constitutions to create such an obligation.) And no state has made pre-K education compulsory for young children, as every state does for those of school age. Pre-K education thus remains for the most part optional from the standpoint of both providers and consumers—an incalculable difference from the primary-secondary arena.

Governance and regulation differ, too. While K–12 education now places heavy emphasis on state-prescribed academic standards, assessments, and outcomes (as well as continued—I would say excessive—regulation of inputs and delivery systems),

pre-K education remains fixated on resources, services, licenses, and credentials.

Most early-education experts are similarly fixated. Indeed, the pre-K policy arena still functions for the most part in a pre-Coleman world of spending levels, staffing ratios, and college degrees rather than cognitive expectations, pupil assessments, and results-based accountability.³¹ Nor are the data on outcomes much good. Even where states try to prescribe the desired results or regulate the curricular content of pre-K programs, except for Florida they end up with scant information on whether that content is being successfully imparted to small children, much less whether it's retained later. The situation is hard to rectify, too, because assessment in this domain is underdeveloped and heavily disputed, because many early-childhood educators care more about non-cognitive elements of child development and because (as in K–12 education) existing providers are loath to be judged by the results of their efforts.

In partial contrast with the K–12 world's familiar distinction between school inputs and student achievement, pre-K analysts tend to distinguish between two broad genres of quality criteria: "structural" quality, which deals primarily with such organizational characteristics of a program—I usually term them "inputs"—as child-staff ratios, class sizes, teacher credentials, and physical safety; and "process" quality, which focuses on interactions among children and staff, environment, and other kids.³² There's a presumption (with good research validating it) that such "process" interactions, when skillfully and regularly deployed in classrooms, are associated with more cognitively developed and school-ready youngsters. As a recent report by The Albert Shanker Institute notes, "Of particular importance is the quality of instruction, which appears to have a vital, lasting effect on building children's cognitive and social skills through the elementary school years."³³

In the present education-policy environment, with its heavy cognitive emphasis and school-readiness focus, it would be valuable to add a third metric for judging quality: pre-K programs' actual results gauged in terms of their "graduates'" readiness for academic success in kindergarten and beyond. Today, however, we're a long way from acceptance of that view across much of the early-childhood community. (Florida is again an exception—but many experts scorn the Sunshine State approach.) In fact, none of the three most widely used sets of "quality" criteria in this field pays much heed to learning outcomes; and by and large, they don't do very well by "process" measures, either. They mostly focus on "structural" (i.e., input) considerations.

For decades, most experts have relied heavily on a metric called the Early Childhood Environment Rating Scale (ECERS), developed by Thelma Harms and the University of North Carolina's Frank Porter Graham Child Development Institute. Its stated purpose is to "assess group programs for children of preschool through kindergarten," and the current version clumps 43 program characteristics into seven categories. A few elements deal with interactions between preschoolers and their teachers but none deals directly with outcomes or school readiness. Most focus on a center's resources, routines, staffing, and activities.

As Abt Associates' Jean Layzer and Barbara Goodson point out, the ECERS scale "was originally developed as a tool that centers could use for self-assessment to target areas for improvement...[and] reflects a generous expansive vision of what is necessary to create a comfortable and nurturing center environment for children." ECERS was not, however, designed with an education-outcomes orientation or kindergarten-readiness emphasis. Indeed, it was not created with any cognitive or curricular focus, and it neglects or undervalues elements of school readiness that matter in kindergarten. "For example," say Layzer and Goodson, "the four

ECERS items that measure the extent to which the environment meets adults' needs made the same contribution to the [center's] total score as the four items that assess the quality of children's language and reasoning experiences. As a result, it is possible [using ECERS] to give a highly favorable rating to programs that offer minimal support for language and literacy acquisition."³⁴

Also quite influential in early-childhood program evaluations and quality judgments are the standards of the National Association for the Education of Young Children (NAEYC), a voluntary accrediting body that has given its stamp of approval to thousands of such centers. For many years, its standards were almost entirely structural. Today, they span ten areas from curriculum to management. Here's a sampler of what NAEYC looks for in programs seeking its approval:

Curriculum: ...Children are provided opportunities to experience oral and written communication in a language their family uses or understands....Children have varied opportunities to develop vocabulary through conversations, experiences, field trips, and books....

Teaching: Teaching staff create and maintain a setting in which children of differing abilities can progress, with guidance, toward increasing levels of autonomy, responsibility, and empathy....Teachers notice patterns in children's challenging behaviors to provide thoughtful, consistent, and individualized responses.

Teachers: All teachers have a minimum of an associate's degree or equivalent. At least 75 percent of teachers have a minimum of a baccalaureate degree....All teaching staff have specialized college-level course work and/or professional development training that prepares them to work with children and families of diverse races, cultures, and languages.

Physical environment: There is a minimum of 35 square feet of usable space per child in each of the primary indoor activity areas....Toilets, drinking water, and hand-washing facilities are within 40 feet of the indoor areas that children use....

Leadership and management: The program administrator responds proactively to changing conditions to enhance program quality....The program administrator and other program leaders systematically support an organizational climate that fosters trust, collaboration, and inclusion.

Such criteria appear wholly worthy on their face, almost unimpeachable (once you get past a touch of political correctness). Who would want a three- or four-year-old child in any other sort of environment? No one credibly claims that environmental and structural considerations are wholly irrelevant, since no one thinks it's good for little kids to be in unsanitary, chilly, or hot places or in the hands of adults with bad habits or police records. It's worth noting, too, that the current NAEYC standards, adopted in 2006, are far more explicit than their predecessors regarding curriculum, assessment, and cognitive development.

Yet even after all the revising and updating, the standards are still chiefly concerned with resources, services, aspirations, and activities. They do not convey any explicit sense of what young children need to learn or how to determine whether they've learned it. Scant attention is paid to process considerations and less to cognitive outcomes. And some of the NAEYC standards' most-cited-and-deferred-to elements—such as pre-K teachers possessing bachelor's degrees—have at best a mixed and nebulous bearing on student learning.³⁵

Learning, however, is not the overriding consideration. One pores in vain through NAEYC's standards in search of any clear statement

that the mission of early education is to prepare children to thrive in kindergarten and beyond. Rather, the stated purpose is to ensure that accredited programs “advance children’s growth,” are “consistently nurturing and filled with learning opportunities,” and so on.

While NAEYC and ECERS focus on individual programs and centers, Pew-supported NIEER concentrates on state policies. It evaluates them against ten “quality standards,” eight of which pertain to resources and services: class-size limits and kid-to-teacher ratios, staff credentials, meals, referral services, and so forth. The two that deal with content and assessment are both sound as far as they go. One says that the “National Education Goals Panel content areas covered by state learning standards for preschool-age children must be comprehensive.” The other says that “Site visits must be used to demonstrate ongoing adherence to state program standards.”

But look again. The site-visit standard says nothing about what those “state program standards” should be and is mute about educational outcomes or how to gauge them. This is classic “compliance” language, not “results” language.

The content-areas standard—invoking the National Education Goals Panel specifications circa 1991—isn’t bad. (Those content areas, by the way, are “children’s physical well-being and motor development, social-emotional development, approaches toward learning, language development, and cognition and general knowledge.”) Note, however, that the text says only that states should have “comprehensive” learning standards in those areas. It doesn’t say what the standards should include or that the state must have a mechanism to determine whether its pre-K programs in fact meet them.

Ironically, the Goals Panel itself recommended almost two decades ago that states adopt comprehensive assessment systems to determine children’s readiness for school.³⁶ But nothing like that has made it into the NIEER policy check-list for state pre-K programs.

Given such flabby criteria for quality, we shouldn't be surprised that relatively few studies of early childhood programs do a great job of appraising their educational efficacy. Many don't even try. In 2006, Martha Zaslow and colleagues examined 65 studies of child-care quality published over the previous quarter century. Barely half of them even focused on "language development" and/or "cognition and general knowledge," and among those that did, more than one-third had methodological shortcomings.³⁷

In short: although it's been trying to make its way into the 21st Century, post-Coleman understanding that education quality is in the end not about what goes into a program or happens there but about the results that are achieved, the pre-K world has not yet reached this destination. It displays widening recognition that its programs and policies *ought* to have a standards-and-results orientation. At the same time, it harbors continuing arguments over what those standards ought to be, how to assess whether they're being met, and whether anyone is or should be held to account for producing success or failure.

Without widespread use of agreed-upon quality metrics that address program outcomes and school readiness, it's no surprise that available data about early-childhood offerings, operators, and participation rates are so murky with regard to "education" versus "child care" as well as program effectiveness and cost compared with benefit. Economists David Blau and Janet Currie, after reviewing much of the relevant literature, found that while "process quality is more closely related to child development than structural [i.e., input] quality...there are no nationally representative data available on process measures. Researchers must rely on structural measures under the assumption that the two types of quality are related. Complicating matters further is the failure of

the U.S. child care data collection system to collect quality data on a regular basis.”³⁸

Other analysts voice similar doubts about the established quality criteria in this field. Robert C. Pianta, dean of the University of Virginia school of education and a developer of assessment methods that have yielded empirical data on thousands of pre-K classrooms, says that what actually makes a difference in pre-school learning is well established but that the usual criteria bear scant relationship to it. “The evidence,” he says, “is quite clear that it is the teacher’s *implementation* of a curriculum, through both social and instructional interactions with children, that produces effects on student learning.” By that he means teachers who “strategically weave instruction into activities that give children choices to explore and play,” including “explicit instruction in certain key skills; responsive feedback; and verbal engagement/stimulation.”³⁹

When Pianta applies that definition of quality to today’s pre-K programs, he doesn’t find much of it. His research team determined that “only about 25 percent of classrooms serving 4-year-olds provided students with the high levels of emotional and instructional support that are needed....Unfortunately, exposure to gap-closing classroom quality, although highly desirable from nearly every perspective imaginable, is not a regular feature of early schooling and even less likely for children in poverty.”⁴⁰

This is not good news for advocates of universal preschooling. What Pianta is saying is that when educationally relevant “process” criteria are systematically deployed in reviewing existing programs, high-quality pre-kindergarten turns out to be a rarity. Worse, the criteria that are more commonly used—the “structural” kind, with their emphasis on inputs and ratios—bear little relation to the research into what constitutes educational effectiveness. Blau and

Currie find “little convincing evidence that structural child care inputs affect child outcomes.”⁴¹ Or, as Pianta explains it:

Many states and localities measure program “quality” only in terms of proxies—the credentials of teachers, the size and spaciousness of the facilities, the amount of learning material available, and the length of the preschool day. Except for the last characteristic, these “quality indicators” do not measure what programs offer young children that is educationally important. Still, these indicators often drive program design and policy.

Pianta and his colleagues mounted an elaborate study to see whether either the ECERS benchmarks or the NIEER program standards were closely linked with the actual classroom behaviors and activities most apt to foster “academic, language and social development” among four-year-olds. The results of their investigation showed almost no relationship between those widely used standards and actual cognitive outcomes for kids. While the authors did not openly criticize the criteria of ECERS and NIEER (and, implicitly, NAEYC), indeed suggested that such criteria might help to screen programs for basic adequacy, they also found that classrooms and centers that fared well on those criteria did not necessarily boost the school readiness of their young participants.⁴² (Pianta and colleagues have developed their own assessment of teacher-child interactions in the pre-K setting, particularly instructional interactions, that does predict gains in kids’ outcomes.)

Why does so much of the pre-K field fuss about inputs and structures when these do not equate to program effectiveness and thus not to quality as properly defined for today’s education world? Because, for one thing, inputs are far easier to measure. The human interactions that truly matter can only be gauged by placing sophisticated

observers inside classrooms over long periods. This is doable—Pianta and colleagues have trained more than a thousand such observers—but it’s time-consuming, relatively costly, and (unless done with care) vulnerable to charges of subjectivity and inconsistency.

Another reason is surely habit, the fact that ECERS and NAEYC were evaluating day care and preschool programs long before today’s preoccupation with kindergarten readiness and gap-closing. NIEER, too, has been making waves for years now. Those organizations’ measures, criteria, and emphases simply cannot bear the burden we’re now asking them to carry, as if we were expecting the sanitation inspector also to appraise the nutritional value of the restaurant’s cuisine. But once something gets used in certain ways and that usage becomes widespread, it attracts believers, practitioners, and habitués, and becomes harder to alter.

Yet another explanation, certainly the most troubling, is that the kinds of quality criteria for pre-K programs that Pianta and others favor typically yield glum findings precisely because so few pre-K classrooms display such attributes as high-quality interactions between skilled teachers and children. Wide application of the proper criteria would lead to many extant programs, centers, and operators being found lacking, a profoundly unwelcome message for advocates and operators.

“If one were to rest the whole system on those structural indicators that people tend to talk about,” Pianta says, “you could vastly overestimate the level of quality that is in the system.”⁴³ This can be discouraging. Pianta himself is pessimistic about large-scale implementation of high-quality programs, noting that “when these approaches are disseminated to large groups of preschool teachers through districtwide training or college courses, such approaches typically have a much-reduced effect on outcomes, often because the quality of implementation is low.”

A further explanation is philosophical resistance to a cognitive view of pre-K quality. Particularly when dealing with small children, adults must certainly attend to “the whole child” and his/her varied developmental needs. Nevertheless, in today’s pre-K policy context, what matters most is a program’s effectiveness in imparting essential school-readiness skills to its young participants, principally in the cognitive domain. Key attributes of such programs include clear goals, accurate assessments, and a willingness to be judged by outcomes, as well as by the high-quality classroom interactions most apt to yield them. But that isn’t how most early-childhood educators prefer to view their work, much less to be evaluated on their performance.

This is a field, like so much of American education, whose culture and belief structure have long been profoundly “child-centered,” “nurturing,” and pedagogically progressive, in love with wild flowers and repelled by crop cultivation. E.D. Hirsch depicts this “thought-world” in his important book, *The Schools We Need and Why We Don’t Have Them*:

From Romanticism, the American educational community inherited the faith that early childhood is a time of innocence and naturalness, a time for *being* a child....It is wrong to spoil the one time of life when children can develop in tune with the order of things....Self-evidently, premature book learning goes against nature. According to the educational community, “research has shown” that untimely interventions and constraints are “developmentally inappropriate” and create a hothouse, forced-feeding atmosphere....Such expert attacks against early book learning intensify the already powerful Romanticism in American culture....⁴⁴

Parents, too, are not of one mind in selecting preschool and day-care arrangements for their children. When asked on a recent survey what factors were “very important” in choosing centers for their (three-through-five-year-old) youngsters, “learning activities” were cited by 81 percent. But 88 percent emphasized “reliability,” 73 percent looked for “time with other children,” and 60 percent were concerned with “location.” (Interestingly, just 37 percent deemed cost “very important,” though that rose to 53 percent among parents below the federal poverty line.)⁴⁵

In the early childhood years in particular, family priorities differ greatly, and particularly because nothing at this age is compulsory, a complex and messy system has arisen that caters to varied tastes and needs. It’s hard to imagine homogenizing such a system—and far from clear that doing so would make sense from a policy standpoint or prove politically acceptable.

What changed over the past quarter century in the compulsory domain of K–12 education was that “outside” reformers—governors, business leaders, even Congress—insisted that schools prove themselves successful according to how much their pupils learn in relation to pre-set standards and learning objectives. That is by no means a flawless strategy and I have often faulted the standards-based reform effort in primary-secondary education for its shortcomings.⁴⁶ But in time we’ll get it close to right, and external accountability for demonstrated results will remain the name of the K–12 game.

Nobody would want a young child’s readiness for kindergarten to be gauged solely in terms of cognitive achievement. Dressing oneself, standing in line, sharing toys, and gaining control over large and small bodily movements are crucial, too, and we know from the NCLB experience the distortions that arise when too

much accountability is tied to narrow measures of performance. But today's pre-K world remains a long way from having to worry about that dilemma. Its current problem in gauging program quality is not just its disinclination to judge children's school readiness but also its aversion to judging programs by their results (however defined and measured) or even by their diligent use of practices that are known to foster good results. It's pre-Coleman, indeed.