

**New York State Universal Prekindergarten
Assessment Meeting**

May 15-16, 2000

**Cornell Early Childhood Program
Department of Human Development
College of Human Ecology, Cornell University
Ithaca, New York, 14853
Telephone: 607-255-2457
E-mail: cecp@cornell.edu**

**Funding Provided by the Foundation for Child Development
and the A. L. Mailman Family Foundation**

**Summary Prepared By Carol H. Ripple
Ripple Consulting**

Table of contents

Acknowledgements	1
Executive Summary	2
Introduction	4
Meeting Structure and Goals	5
What is School Readiness?	6
Moving Toward Evaluation: Conceptual Considerations	11
The Case In Point: UPK Evaluation Implementation	16
Next Steps	21
Postscript	22
List of Participants	23

Acknowledgements

This report and the working conference that generated the ideas it contains could not have been possible without the generous contributions of many wonderful individuals and organizations. Appreciation is expressed first to the Foundation for Child Development and the A. L. Mailman Family Foundation for providing the financial support needed to hold the conference and prepare the report. Special thanks go to Ruby Takanishi and Fasaha Traylor of the Foundation for Child Development and Luba Lynch of the A.L. Mailman Family Foundation for the confidence they have shown in this project as part of the ongoing work of the Cornell Early Childhood Program.

All conference participants contributed greatly to a very stimulating and productive set of discussions. Particular gratitude goes to the UPK program directors and teachers in attendance for keeping the group grounded in the real world of children, families, and communities. Denise Gomer and Diana Levy of the Ithaca City School District were particularly helpful in assisting us to think through the conference agenda and identify participants from the Ithaca community.

Cynthia Gallagher, Team Leader, Early Childhood Coordination Team at the New York State Education Department and prime mover behind the state-wide Universal Prekindergarten initiative, supported the idea of a conference to discuss assessment of UPK impacts from the beginning and was instrumental in identifying key participants. Eleanor Ukoli, Director of the Office of Early Education at the New York City Board of Education, was also supportive and made sure that New York City perspectives were included in our discussions.

A note of appreciation is extended to Gary Henry, Kim Kopko, Kristi Lekies, John Love, Judy Ross-Bernstein, Elizabeth Stilwell, and Fasaha Traylor for the careful reading given to and the helpful feedback provided on an earlier draft of this report. Carol Ripple deserves special kudos for the outstanding job done in capturing the richness and complexity of conference discussions in such readable form.

Elizabeth Boxer managed all of the logistical details surrounding the conference with efficiency, enthusiasm, and a readily accessible sense of humor. For all those qualities and many more we are especially grateful to Liz. Sarah Lane, Kim Kopko and Susan Werner also pitched in with valuable skills to make the conference proceed more smoothly.

Thanks go as well to Marge Dill for her fine job of facilitating the conference process and keeping us all on task.

Finally, appreciation is expressed to the College of Human Ecology and the Department of Human Development at Cornell for providing space and various other kinds of in-kind support for the conference.

Mon Cochran, Director
Cornell Early Childhood Program

Executive Summary

In the summer of 1997, the New York legislature authorized implementation of the New York Universal Prekindergarten program (UPK). The program began serving four-year-olds in selected applicant districts across the state in the fall of 1998; by 2002, the program is slated to become fully universal, accommodating all eligible four-year-olds in the state whose parents wish them to attend.

After two years, the program is now well underway. With increased operational stability, UPK administrators face growing pressure to heed the call for an evaluation of program effectiveness. Although initial legislation called for an evaluation, the mandate has not yet been funded. Seeking to act before the pressure becomes too great, thereby avoiding the hasty implementation of an assessment strategy, UPK administrators have been considering how best to produce findings that would convincingly answer the question, “Does it work?”

With support from the A. L. Mailman Family Foundation and the Foundation for Child Development, representatives from the field of evaluation research, private foundations, and UPK providers gathered at Cornell University to discuss assessment strategies that might be appropriate for UPK. Specifically, participants discussed the advisability and feasibility of using child outcomes as an assessment mode. In large part, this focus reflects the recognition that the evaluation had to be methodologically rigorous enough to convince state legislators and other stakeholders of UPK effectiveness (or lack thereof).

The gathering was structured to achieve two goals: to present and elicit information that could guide decision-making regarding an outcomes-based evaluation (namely, *if* and *how* child assessment was appropriate); and, if appropriate, to engage providers in conceptualizing a workable statewide evaluation scheme. Early in the meeting, the group heard presentations from two participants with extensive experience in large-scale evaluation research, John Love, from Mathematica Policy Research, and Gary Henry, from the Andrew Young School of Policy Studies at Georgia State University.

Following the presentations, the group engaged in prolonged discussion aimed at developing consensus regarding the definition of ‘school readiness.’ From the outset, it was clear that the term was a source of discontent and concern. The group was able to endorse a broad conceptualization that encompassed five domains: physical well-being; social and emotional development; approaches toward learning; language development; and cognition and general knowledge.

Placing children in the broader ecological context including family, school, neighborhood, and community was also embraced. Just as importantly, readiness was defined not as a child characteristic, but as a relationship between the child and the school context. The role of communities in supporting school readiness was also emphasized. In keeping with a comprehensive, contextual definition of readiness, *everybody* should be ready, not just the children. The concept of readiness needs to move beyond transitioning children into school and into fostering continuity across development.

Despite the group's growing ease with 'school readiness' when it was defined in such broad terms, some discomfort remained. This unease was alleviated considerably when a workable alternative was proposed: the goal of UPK evaluation should be to assess *indicators of child well-being*, rather than school readiness *per se*.

Following the discussion around defining school readiness, the meeting turned to conceptual and practical considerations in evaluation design and implementation. Issues addressed by the group included assessment strategies, sampling (who should be assessed?), design (should there be a comparison group?), measurement, and interpretation and dissemination of results.

The group agreed on several key points. Relying on the standard 4th grade testing schedule was insufficient to assess UPK success. Beyond that, most participants could only consider using child outcomes in evaluation if there was ultimately much more to the overall evaluation that captured the broader context, including schools' readiness for children. The need to produce data that will satisfy stakeholders' need to know about program effectiveness should be balanced with educating them as to what is important to know about children's well-being.

Specific concerns were voiced regarding the nature of the assessment strategy. Only developmentally appropriate, minimally disruptive assessments would be acceptable. Anecdotal measures were endorsed, emphasizing the importance of parent reports of child well-being. There was some acceptance of pre-and-post assessments, of the importance of continuity of measurement over time, of the need for a non-UPK comparison group (acknowledging the challenges of finding one), and of the need for consistency of measurement tools across districts (with flexibility for districts to administer their own measures as well). The difficulty of distinguishing the effects of UPK from those of other preschool program inputs experienced by the same children was identified as a significant problem.

The group ended by developing a list of next steps in the process toward UPK evaluation. A Steering Committee comprised of representatives should be convened to develop recommendations for the process, which would then be disseminated to regional groups. Overshadowing all, however, is the current lack of state funding for an evaluation. Strategies for stimulating this funding were suggested.

In closing, meeting organizer and host Mon Cochran summed up by citing the major challenge ahead. The evaluation of UPK promises to be scientifically complicated, and logistically it is an enormous task. The group was charged with being realistic about goals, design, and expectations, but not to back away from the ideals set forth early in the meeting. The meeting was closed with a pledge to keep all those present involved in the process.

Introduction

Legislation authorizing implementation of the New York Universal Prekindergarten program was passed in August of 1997. Administered by the New York State Education Department (NYSED), the program began in the 1998-99 school year. Current legislation provides three years beyond the 1998-99 school year to reach the goal of accommodating all eligible four-year-olds in the state of New York whose parents wish their children to attend.

New York's program (UPK) is the nation's second state-funded universal prekindergarten program, after Georgia's, to be introduced into legislation. By design, UPK hinges on collaborative arrangements to deliver services. While funds are distributed to qualifying applicant school districts, the legislation requires that at least 10% of district UPK budgets be used to contract with non-public school providers.

Recognizing the need for accountability, the initial UPK legislation called for a program evaluation. The mandate, however, has not yet been funded; to date, all of the operating budget has been needed for program operations. Whereas several districts have established their own assessment plans, these alone cannot answer the need for state-level evaluation.

With support from the A. L. Mailman Family Foundation and the Foundation for Child Development, representatives from the fields of assessment, private foundations, and UPK providers gathered at Cornell University on May 15 & 16, 2000, to discuss assessment strategies that might be appropriate for UPK.¹ Rather than waiting until legislative pressure forced the state UPK program into carrying forward a hastily conceived and implemented assessment design, the meeting was held as a proactive response to mounting political interest in program outcome data.

As introduced by meeting host and organizer Mon Cochran, Director of the Cornell Early Childhood Program, the meeting was intended to identify issues associated with the feasibility and appropriate means of using *child outcomes* as one method to assess UPK. Although child outcomes represent only one possible mode of program assessment, and others might be introduced in other phases of an evaluation, this meeting was designed expressly to address this particular means of measuring UPK's effectiveness. In large part, this focus was in response to growing pressure from several different quarters for data that would be recognized as methodologically rigorous enough to convince state legislators and others interested in the program of its effectiveness (or lack thereof).

Several studies that focus on the nature of UPK and the way it is being implemented are currently under way in school districts across the state. For example, one investigation by Cornell researchers consists of a survey and four district case studies that are documenting what programs look like in the "Wave I" school districts and probing the rationale behind districts' particular implementation schemes. So far, researchers have generally avoided questions of child outcomes, waiting until providers have fully implemented UPK.

¹See Appendix 1 for full list of participants.

Why Now?

UPK Coordinator Cynthia Gallagher summarized a range of influences and circumstances that make this the right time to address UPK program evaluation:

- There is a high level of interest in UPK assessment among policymakers, providers and administrators;
- After 2 years of intense work on program implementation, UPK is ready for a broad-based assessment;
- If we act quickly, we can design a more rigorous assessment that includes a non-UPK comparison group; as of 2002-03, this will not be possible;
- UPK assessment has national importance as other states look to NY for direction in assessing their own prekindergarten programs;
- Although currently unfunded, legislation mandates UPK evaluation;
- As revealed by recent Comptroller's UPK audits, extant data are insufficient to tell us whether the program works;
- School Accountability for Student Success (SASS) emphasizes accountability in NY's education system; UPK should be a part of this effort.

Progress toward conceptualizing a UPK outcome evaluation began in January and February of 2000, when State UPK Coordinator Cynthia Gallagher hosted a series of meetings with the researchers currently launching studies. From those meetings came a better understanding of projects already under way and a renewed sense of the importance of anticipating the pressure to evaluate UPK based on child outcomes. By March, the need for a unified approach to the issue was clear, and this May gathering was the direct result.

Meeting Structure and Goals

Because this meeting was held before pressure from the legislature became too great, the group was able to focus on the question of suitable evaluation strategies: Is it appropriate to base a UPK evaluation on child outcomes? By asking first *whether*, and not jumping ahead to decide *how*, the UPK program community hoped to assume a substantial role in determining the contours of the assessment strategy. In this way, those most directly involved in UPK service provision—the providers themselves—could assume ownership of the assessment plan.

The gathering, then, was structured to achieve two goals: to present and elicit information that could guide decision-making regarding an outcomes-based evaluation (namely, *if* and *how* child assessment was appropriate); and to engage providers in conceptualizing a workable statewide evaluation scheme.

The meeting began with presentations from two participants with extensive experience in large-scale evaluation research. John Love, Senior Researcher at Mathematica Policy Research,

Inc., presented his framework for assessing school-readiness outcomes, and Gary Henry, from the Andrew Young School of Policy Studies at Georgia State University, discussed his work on the Georgia Universal Prekindergarten program impact evaluation. Insights gained from these initial presentations were echoed by the group throughout the meeting. They formed the basis for fruitful discussions that proved essential to the tasks at hand. The remaining time was spent sharing ideas, voicing concerns, and working toward consensus on key issues.

This summary recounts the meeting proceedings. Based on the group's assumption that any UPK evaluation would have to in part address the promotion of school readiness, the first section describes the extensive discussion surrounding this term. Then issues posed by evaluation are summarized. The critical role of context in assessing children's school readiness, and the concomitant importance of placing child outcomes within an ecological model, gained major importance in the discussion. The final sections of this summary document the group's application of these issues to the beginnings of an evaluative framework for UPK.

What Is School Readiness?

Designing an assessment strategy for New York UPK is made all the more interesting because the program goals are not specifically stated in legislation, nor have they been articulated in the regulations promulgated by the State Education Department. Because of this, assumptions need to be made about what the program's goals are, and about what kinds of data are needed to be able to demonstrate program effectiveness. The implementation of UPK through school districts and as part of a larger K-12 initiative places education at the heart of the program's method and goals.² The first assumption of program goals, then, has to do with school readiness. A central goal of UPK as a preschool program is to prepare children to enter kindergarten ready to succeed.

Prior to the meeting, and in preparation for the discussion of school readiness, the group had read a paper by John Love presenting his assessment framework in the context of a national Head Start evaluation.³ Defining the term 'school readiness' itself was the source of considerable discussion that centered around several key issues, briefly described below. Beyond the expressed frustration with the term, there was widespread endorsement of the tenets reflected in Love's paper.

² Universal Prekindergarten was introduced into law in New York as part of the LADDER plan (Learning, Achieving, Developing by Directing Education Resources), a five year, multi-billion dollar initiative put forward by the NYS Assembly that included expansion of full-day kindergarten, reduced class size in grades K-3, improved infrastructure, new school construction, increased aid for textbooks and computer technology, and increased professional development for teachers in addition to universal prekindergarten.

³Love, J. M. (1999). A framework for assessing school readiness outcomes in the context of a national Head Start impact evaluation. Paper prepared for the Advisory Committee on Head Start Research and Evaluation, Mathematica Policy Research, Inc., Princeton, NJ.

Child Readiness Is a Comprehensive Construct

In part, school readiness is what children know, and what they can do. Just as important, readiness consists of those things that are needed to support child development: community support, high-quality preschools, nutrition and health, parent involvement, and parental access to training and support.

Although cognitive development is foremost in some definitions of school readiness, teachers tend to emphasize non-cognitive factors such as physical health and nutrition; the ability to communicate needs, wants, and thoughts verbally; and curiosity.⁴ Reflecting this broad conceptualization, Love described the five distinct dimensions of early learning, development, and abilities that form the core of a comprehensive definition of school readiness, as defined by the Goal One Technical Planning Group of the National Education Goals Panel:⁵

1. Physical well-being;
2. Social and emotional development;
3. Approaches toward learning;
4. Language development;
5. Cognition and general knowledge.

Assessing child readiness should tap all of these dimensions and account for the conditions that support each domain.⁶

However broad its definition, readiness should be an elastic construct that accounts for the vast variability in children's backgrounds and abilities. In a 1995 position statement on school readiness, the National Association for the Education of Young Children endorsed a commitment to achieving readiness for all children that called for: addressing inequities early in life; recognition of and support for individual differences; and the establishment of reasonable and appropriate expectations of children's capabilities on school entry. Because all five-year-olds are not alike, the definition must embrace diversity.⁷

⁴Heaviside, S., and Farris E. (1993). Public school kindergarten teachers' views on children's readiness for school (NCES Report No. 93410). Washington, DC: U.S. Government Printing Office.

⁵Kagan, S. L., Moore, E., and Bredekamp, S. (1995). Reconsidering children's early development and learning: Toward common views and vocabulary. Washington, DC: National Education Goals Panel.

⁶Love, J. (July, 1999) A framework for assessing school readiness outcomes in the context of a national Head Start impact evaluation. Paper prepared for the Advisory Committee on Head Start Research and Evaluation, Mathematica Policy Research, Inc., Princeton, NJ.

⁷National Association for the Education of Young Children (1995). NAEYC Position Statement on School Readiness. Washington, DC: Author.

Not a Characteristic, But a Relationship

Ideally construed, readiness is not a child characteristic. Rather, it is a relationship between the child and the school context that reflects children's level of development and learning that are appropriate for success in school. In the context of multiple domains relevant to both child and school, readiness exists along a continuum: it is not an 'all or none' construct.

Beyond the Child: Community and Family Readiness

Ideally, the optimal interpretation of school readiness data should focus on the collective status of young children as they enter kindergarten. But how do communities support readiness? As formulated by the Goal One Technical Planning Group,⁸ communities can:

- Provide access to high-quality, developmentally appropriate preschool programs.
- Ensure that children receive nutrition and health care.
- Support parents as children's first teachers.
- Provide parents with access to training and support that they need.

What is the level of community-wide readiness? How does it vary by important community subgroups? Asking these questions has practical value as the answers can be applied to mount continuous community improvement efforts that support readiness.

Coming to Terms with 'School Readiness'

From the outset, it was clear that the term 'school readiness' was the focus of discontent and that it required extensive discussion if the group was to agree on its meaning. Love's broad conceptualization resonated with the group, and there was expressed endorsement of multiple aspects of readiness that encompassed and expanded on Love's points.

Beyond endorsing a comprehensive definition, establishing the ecological context for children's readiness was deemed to be essential. This context includes teachers, schools, parents, communities and the ways these environmental systems interact with one another. Readiness must be considered at each of these levels. Even further, the concept of 'legislator readiness' was suggested. One role of the evaluation process should be to raise legislators'

⁸Kagan, S. L., Moore, E., and Bredekamp, S. (1995). Reconsidering children's early development and learning: Toward common views and vocabulary. Washington, DC: National Education Goals Panel.

awareness of their own expectations for the program (and for children), and to educate them as to what is reasonable to expect.

Following are several points raised in the extensive discussion surrounding readiness:

- The word ‘readiness’ seems antiquated, but in some school districts it is the only language they know.
- Getting children ready for school is strictly a political notion.
- The word ‘readiness’ might be keeping us in a conceptual box.
- Child readiness has to do with social experience, problem solving, conflict resolution, exposure to early literacy, science, math, and enthusiasm for learning. It is not skill and drill, but more an issue of enthusiasm for learning and life.
- Children’s ability to communicate effectively with teachers and peers is central to readiness. Pre-academic attitudes, including motivation and curiosity, are important. But children must also possess the skills to communicate these positive traits and attitudes to their teachers. Social and communication skills with peers and adults are important to classroom adjustment, as demonstrated by the fact that weak and ineffective communication skills are associated with behavior problems in four- and five-year olds.
- Readiness has to do with children’s social maturity, and their ability to separate from their parents. Expanding on this, physical, social, and emotional ‘stamina’ is part of readiness. In turn, parental and family readiness influence their children; parents should be acknowledged as children’s first teachers.
- Ready for *what*? Why do we not give validity to where children are at a given moment? When we focus so much on the next step we may be failing to cherish early childhood. We should be giving kids credit for where they are, without pressure.
- The operative word is ‘school’: Is the school ready for the child? Are teachers using developmentally appropriate practices and appropriate curricula? Are teachers trained adequately, and are class sizes small enough? Moving children from small prekindergarten to large kindergarten classes limits their progress.
- Whereas we want a core consensus, readiness is a community-level process that may vary from district to district. There may be specific expectations that are couched in individual, distinct community contexts. Some districts have already engaged in assessing children’s needs as part of their implementation process. Encouraging districts to articulate their definitions of readiness might yield fruitful discussions that could develop into local evaluation interests.

- What are the readiness expectations, and what is the basis for setting them? What is needed to prepare children for success in school?
- What are the conditions that support readiness? What is the level of social readiness? Is society ready to commit the necessary resources?
- Different constituencies have different conceptualizations of readiness (e.g., kindergarten teachers, preschool teachers, and parents). We need to find out what these are and engage in a dialogue that moves us closer to a fuller understanding.
- *Everybody* should be ready: Readiness is a set of expectations aimed at children, but it needs to move beyond that, beyond transition and into *continuity*.
- Readiness is the relationship between where children are and where they need to go next.

Introducing the concept of variability in readiness according to contextual factors eased fears over adopting a rigid definition that might be ultimately untenable. Whereas school districts are generally driven by test scores, a broad conception of readiness encompasses multiple domains within and beyond the child. By capturing the variability across domains, and getting beyond test scores, the evaluation should avoid the pitfalls associated with narrowly defined outcomes.

An Agreeable Concept

Despite the group's growing ease with 'school readiness' defined in broad terms, some discomfort remained. The unease was alleviated considerably when a workable alternative was proposed: the goal of UPK evaluation should be to assess *indicators of child well-being*, rather than school readiness *per se*. This term, it was felt, reflects an appreciation of UPK as support for children, rather than as intervention. It captures school readiness as a relationship (i.e., between child and expectations) better than 'school readiness' does and conveys greater sensitivity to individual differences.

In sum, the group expressed cautious endorsement for a multi-dimensional, multi-domain approach to indicators of child well-being as a working definition of school readiness. Assessments, then, would measure the five dimensions identified by Love, and would place child readiness in an ecological framework that includes parent, teacher, school, system, and community readiness.

Moving Toward Evaluation: Conceptual Considerations

Collecting child outcome data to assess school readiness is both appropriate and feasible. However, whether it should be done or not is another matter, and the decision rests at least in part on the ideal assessment scenario, the practical realities, and the relationship between the two (how do we get from ideal to real?).

--John Love

The next two sections document the shift from defining school readiness to discussing conceptual and practical considerations in evaluation. In the present section, background information on evaluation from the presentations by John Love and Gary Henry are highlighted.

Assessment Strategies

Ideally, assessment strategies incorporate multiple modes of assessment as well as multiple perspectives on outcomes. Modes include direct child assessment, child ratings (e.g., by teachers), and classroom observations. Perspectives may be gathered from teachers, parents, the children themselves, and others.

Timing of assessment is also crucial. Should data be gathered in the first and second grades? In the Georgia evaluation, research demonstrating the mediating effect of kindergarten quality on Head Start children's later outcomes⁹ was used to guide the decision to assess children beyond entry into kindergarten.

Sampling: Who should be assessed? An essential part of an assessment strategy is deciding who to include in the evaluation. Particularly in a large-scale evaluation, assessing all children in the program is neither practical nor advisable. Determining which children to assess, then, is one of the first challenges in implementing an assessment strategy. Sampling procedures should ensure that the children sampled represent the diversity of the entire prekindergarten population. Depending on the nature of the selected measures, smaller subsamples of participants may be selected for more intensive, costly assessment strategies (e.g., classroom observations), whereas less costly measurement schemes can be used for the larger sample.

Sampling procedures are determined in large part by decisions regarding the sampling unit that will be represented in the results. In the Georgia example, counties and cities were selected as sampling units.

⁹Lee, V. E., and Loeb, S. (1995). Where do Head Start attendees end up? One reason why preschool effects fade out. Educational Evaluation & Policy Analysis, 17, 62-82.

Program Evaluation: A Primer

Generally speaking, program evaluations are performed for the following purposes:

- Assessment of program merit and worth (Is the program effective? For whom, under what circumstances?).
- Program or organizational improvement (What are the strengths and weaknesses in the program?).
- Oversight and compliance (Does the program meet formal guidelines and requirements?).
- Knowledge development (Why does the program work?).

These purposes can be met by conducting the following types of inquiries:

- Descriptive, or process (client and program characteristics, program resources, and expected outcomes).
- Classification (Develop taxonomies, groupings, or categories across clientele, services, programs, settings, effects, etc).
- Causal analyses (program effects, causal mechanisms; involves a comparison group).
- Values inquiry (underlying assumptions and inherent limitations).

Steps to implementing an evaluation are:

- Determining processes for deciding on the outcomes.
- Selecting the measures.
- Deciding on assessment procedures.
- Determining how to sample children.
- Deciding on levels of aggregation for analysis.
- Analyzing and interpreting data.
- Providing reports to stakeholders.
- Creating the feedback loop.

Source: Gary Henry, Georgia State University

The comparison group dilemma. In addition to determining what portion of the prekindergarten population to assess, a major issue involves whether or not to include a non-program comparison group. Without a comparison group, it is hard to produce outcome findings that convincingly demonstrate that a program benefits children above and beyond what they could obtain elsewhere, and that participants do better on chosen outcomes than children who do not receive the program. Although it is possible to compare data to norms when using well-established assessment measures, this is not a substitute for a carefully selected comparison group of children that closely resemble those who are in the program.

Whereas performing a high-quality evaluation of children who are participating in a program can be quite challenging, the identification and inclusion of a suitable comparison group

is often the most difficult part of program evaluation research. A number of complicating factors, such as participant recruitment, accounting for variables that determine who is in programs and who is not, and fully documenting the experiences of non-program children, make selection of an appropriate comparison group a challenging process. The dearth of children who are true “control” cases (i.e., who are not receiving any services at all) requires careful sampling and extensive data collection procedures. These issues are compounded in New York because of UPK’s universal intent. Upon full implementation in 2002-03, finding unserved children should, by definition, be nearly impossible to do.

Disentangling the Effects of UPK from the Effects of other Programs. Another concern flowed from the fact that NY UPK is only a 2.5 hour program, so that children in full-day child care arrangements experience it as part of a broader program or array of programs. Is there research evidence that a 2.5 hour preschool program can have positive effects independent of other early childhood programs of longer duration, in which the same children participate? In these (very frequent) instances, how can the effects of full-day child care or Head Start, for instance, be distinguished from those of UPK?

Measurement considerations. Practical considerations pose challenges to valid measurement. The biggest challenge may lie in designing culturally, racially, ethnically, and linguistically appropriate procedures. Measures also should balance positive and negative indicators, adapt to local circumstances, and be flexible and expandable.

Georgia’s Universal Prekindergarten: Are there lessons for UPK?

Gary Henry presented his experiences with evaluating Georgia’s universal prekindergarten program, the nation’s only other state-funded universal prekindergarten program.

The program. Funded by dedicated state lottery proceeds, Georgia’s program enjoys considerable political popularity and expanding allocations (\$216 million in fiscal year 1999, serving 60,471 children). Georgia prekindergarten is implemented statewide; however, some counties have such low populations of eligible children that they do not offer the program. More than half of the children enrolled in the program live within the Atlanta city limits. The 6.5 hour/day program is delivered by three auspices: public schools, non-profit child care, and for-profit child care.

The evaluation. The Georgia assessment is now in its fifth year of a 12-year study, and evaluators are recommending that they shut down the initial phase and start on new cohort of children and families. The current cohort consists of approximately 3,000 children in over 1,800 classrooms. Not all children are represented on each outcome; on any given assessment, from 750 to 1000 children provide data. There is no comparison group.

Because developing new measures is time-consuming and costly, using existing measures rather than developing new ones can ease assessment costs and processes. Further, using established, norm-referenced measures provides *de facto* comparison data.

Gary Henry emphasized the importance of focusing the evaluation results on *indicators* instead of definitive measures in the Georgia evaluation. In Henry's experience, presenting results that consist of levels of an array of indicators has been successful; of the wide range of possible indicators (e.g., grade-level retention/promotion, absenteeism, health status, level of parent involvement in grades K-2, etc.), different stakeholders were likely to be interested in different ones. As John Love pointed out, the five dimensions of child readiness support a broad range of indicators. Following presentation of indicator data, analyses designed to investigate what is behind them (i.e., causal mechanisms, such as community characteristics or preschool quality, for example) might be pursued as a secondary focus.

Administrative feasibility. Large-scale program evaluations are typically performed by a single contractor or university which develops local research partnerships to facilitate data collection. In addition to feasibility issues, using local researchers makes it more likely that those collecting the data will have better knowledge of the local context. Further, local researchers can adapt data collection to include measures of interest to specific providers or districts to add to the basic "common core" statewide protocol.

Interpretation and Dissemination of Findings

Once the data have been gathered and analyzed, two other steps remain that are equally important to successful evaluation: interpretation and dissemination.

Interpreting the results. How should evaluation data be construed? Sensitivity to the appropriate level of reporting is critical to how the data are understood. For example, findings could be summarized by district, community, classroom, child, or statewide.

Program auspice presents another consideration to interpreting findings that is particularly salient in UPK. What are the implications of the various auspices of UPK operation (public schools, Head Start, and a range of child care providers) for defining the unit of measurement? If one specific goal of the evaluation is to learn about distinct program effects by auspice, then the design will differ from an evaluation that aggregates data across programs to examine whether or not there is an overall program effect. Part of this decision rests on the expectations for UPK as a whole. This issue also has implications for choosing outcomes: if different programs have different expectations, is uniform measurement still possible?

Ideally, the following guidelines should be taken into account:¹⁰

- Report data on the collective status of entering kindergartners;
- Obtain community-wide estimates of child well-being;
- Estimate child well-being for all of the important subgroups within the community;
- Use findings for continuous community improvement efforts;
- Tie outcomes to supporting conditions;
- Evaluate the implications of the findings.

Matters surrounding interpretation of evaluation findings are critical to safeguarding against their inappropriate use.

Disseminating the findings. Once the data are collected and analyzed, how should they be disseminated to the legislature and to the public? Whereas there is no one model for dissemination that is likely to fit all efforts, the group was interested to hear about Gary Henry's experience in Georgia. He explained that the evaluation findings had to be made very public, and that he made extensive use of the media to accomplish this. He met with the education reporters to explain what the findings meant, to the point of walking them through some results item by item. Henry's careful use of the media was particularly important because, in his experience, legislators were more likely to read the newspapers than they were to read research reports on the findings.¹¹

The Ecological Framework: Assessing Context

In the end, assessing children's levels of school readiness implicitly evaluates the preschool experience. However, it is important to bear in mind that there is a host of experiences that bear on child development during these formative years that ensure a diverse range of experiences across children. This makes it challenging at best to sort out the effects of prekindergarten from myriad other influences and requires the commitment of the community to engage in the process.

Practical and Other Considerations

Ideal frameworks for assessing school readiness are, of course, tempered by practical considerations. Is the community willing to engage in and commit to assessment? Will parents and teachers give their time to the process? How will the efforts be coordinated and supervised, and how will quality in methodology be assured? How will the data be managed?

¹⁰Love, J. M., Aber, J L., and Brooks-Gunn, J. (1994). Strategies for assessing community progress toward achieving the first national educational goal. Princeton, NJ: Mathematica Policy Research, Inc.

¹¹Reports on the findings are available online at the Applied Research Center at Georgia State University site, <http://www.arc.gsu.edu/prek/project.htm>

The Case In Point: UPK Evaluation Implementation

A number of currently ongoing process evaluations¹² will provide essential data on UPK programmatic approaches. UPK's structure ensures a wide variety of implementation strategies; thus, there is likely to be variability between programs in domains such as staff contributions to program operations and decision making, staff training and responsibilities, budgetary allotments, parents' roles, how programs work with other community agencies, and the existence and implementation of a particular curriculum.¹³ These variations must be fully understood and taken into account when considering the design of an outcomes-based evaluation and interpretation of its findings.

Politics and Evaluation

At the heart of political pressure to produce outcome data on UPK is legislators' desire for assurance that the investment in UPK is sound. Providing this assurance must be part of what drives a UPK evaluation: ideally, data should show that children are functioning well and that UPK has made a difference. Producing broad program information that shows UPK is meeting all of the program requirements will not suffice because such findings do not show success in terms of outcomes. In addition to requiring outcome data, legislators are likely to insist on a comparison group.

Having acknowledged the need to produce data to satisfy legislators, the group went on to express interest in making the evaluation process an educational one. One approach could be to ensure that they understand what the baseline is by reporting on children's pre-program levels of nutrition, health, and parental support. These data could create the opportunity to educate legislators and other stakeholders about the importance of multiple domains of child well-being.

A second goal of this approach to educating legislators lies in debunking the inoculation model. This once-popular model rests on the assumption that if intervention is successful it will 'fix' the child for life. Implementing a broad-based assessment that endorses an ecological model provides a means to challenge the deeply entrenched expectations that any single intervention can solve all difficulties.

¹²For example, see Hicks, S.A., Lekies, K. S., and Cochran, M. (1999, June). Promising practices: New York State Universal Prekindergarten (Expanded edition). Ithaca, NY: The Cornell Early Childhood Program.

¹³The national evaluation of the Early Head Start program is similarly challenged to carefully document individual program implementation and practices. These process data will figure prominently in the analysis and interpretation of outcome data. For details see Kisker, E. E., Love, J. M., Raikes, H., Boller, K., Paulsell, D., Rosenberg, L., Coolahan, K, and Berlin, L. Leading the Way: Characteristics and Early Experiences of Selected First-Wave Early Head Start Programs. Volume I: Cross-Site Perspectives. Report submitted to the Administration on Children, Youth and Families. Princeton, NJ: Mathematica Policy Research.

UPK Program Implementation Highlights

Several aspects of New York State UPK implementation will strongly affect evaluation structure. Some that are particularly relevant to evaluation include:

- School districts are required to use at least 10% of UPK funds for implementation in non-school-based settings, including Head Start and child care. Each auspice has its own pre-existing structure and approach to service delivery. From this standpoint, UPK may be seen as a funding stream rather than an independently existing model or approach.
- By 2002-03, more than half a billion dollars will have been invested in UPK. All \$500 million/year goes to districts for program delivery; if that amount is cut then the program suffers (in other words, there is no existing state funding for evaluation).
- New York State encompasses three very different contexts: New York City, the upstate urban centers, and the largely rural remainder of the state. UPK must be structured to suit each of these different contexts.
- UPK was implemented to operate alongside the pre-existing New York Experimental Prekindergarten Program (EPK). How and when to merge these two large state preschool programs is currently under discussion.
- Legislation calls for full, statewide program implementation in 2001-02. Because by then all eligible children should have access to UPK, this has implications for program evaluation. Recruiting a suitable comparison group of children after that point will be extremely difficult.
- In the 1999-2000 school year about 47,000 of the 200,000 four-year-olds in New York state were served by either UPK or EPK. UPK is being phased in by district and by expanding numbers served within districts over time. Districts are not mandated to implement UPK; most districts that have not yet applied for UPK funds are relatively small.
- UPK typically operates for 2.5 hours/day for the school year, but is often delivered as part of a full day, full year program.

Is It Appropriate To Evaluate UPK?

When asked whether or not it is appropriate to use child outcomes to evaluate UPK, the group expressed a wide range of opinions. There was agreement that relying on the standard fourth grade testing schedule was insufficient to assess UPK success, not only because of its narrow focus but also because the time between UPK attendance and fourth grade is unreasonably long. Most participants could only consider using child outcomes in evaluation knowing that there would be much more to the overall evaluation than child outcomes alone.

In general, the group expressed the following thoughts:

- Measuring child outcomes is acceptable, but assessments must be relevant to the program the child is in and to stakeholders. Performing assessments depends on stakeholders' willingness to participate.
- Only developmentally appropriate methods may be used, preferably using existing measures that are least intrusive. There was some acceptance of pre-and-post assessments, and of the importance of continuity of measurement over time.
- Use of anecdotal measures was mentioned several times (e.g., parent and teacher reports on changes in child well-being, tracking growth over time).
- There was a range of opinion regarding who should perform observations, and over what period of time.
- Several participants reiterated the endorsement of measuring child well-being that, at least conceptually, takes the burden off of individual children.

To be successful, assessments must be relevant and acceptable to children, parents, providers, policymakers, and researchers. Pencil-and-paper tools have negative connotations in the field of early care and education, whereas anecdotal measures are perceived as having greater validity. Portfolios of children's work are popular, but perhaps the most important point is not to put pressure on children in the course of assessment.¹⁴

Unit of Measurement: Will Classroom-Level Data Suffice?

The group agreed that child outcomes must be only one *part* of the overall evaluation plan. Participants recognized, however, that it is perhaps the hardest to design and carry out without pressuring children. Concerns over gathering data that could be linked to individual children's performance, and potentially misused as the basis for individual assessment, were also voiced. Based on literature linking classroom quality with enhanced child well-being, members of the group wondered whether it might be enough to measure classroom effects and avoid individual assessment.

Based on his experience in Georgia, Gary Henry indicated that the bottom line for legislators has to do with what is happening with children: Are they better off? In addition, although classroom-level data could consist of counts of problem behaviors, for example, these data would not account for variations in classrooms, year to year and class to class, where there is naturally high variability in the number of disruptive children. Classrooms can be assessed, but this level of data cannot substitute for individual data.

¹⁴This point was widely acknowledged as critical to a successful assessment plan not only because of the importance of protecting children in the process, but also because pressuring children will make both teachers and parents angry and protective.

Should Common Assessment Tools Be Used in All Districts?

Several districts currently have multiple assessment tools in place, but there is no systematic effort to look at data across programs. There was consensus that whereas different districts may emphasize different domains in their programs, agreement on core set of indicators should ultimately be reached. These indicators, then, would be measured statewide and would form the basis for the evaluation protocol.

From a scientific point of view, these indicators should be measured with the same instruments across all programs. The group supported this concept, acknowledging that it is hard to make a persuasive statement about broad-based program effects without common measures. The assessment should include measures that capture the five dimensions identified by the Goal One Technical Planning Group, as described by Love in his presentation. In addition to common measures, assessments should occur at roughly the same time interval across districts.

Because some districts have already implemented some form of assessment, they should be encouraged to keep their own system in place if they choose. Each district could then add their own set of indicator measures to the statewide protocol, resulting in the potential to include two sets of measures: one common across all districts, and a second that is specific to individual districts.

Some specific ideas on evaluation content and process were offered:

- Parent data should be used: they provide an articulate profile of how children are doing over the preschool year. A parent observation form is critical.
- Literacy environments enhance children's well-being; assessing concepts of print could measure this. Even if teachers 'teach to the test' in this case, it is not a problem because the measure taps important skills that transcend the measure.
- Draw-A-Person was suggested as a way to tap development without creating a negative 'teaching to the test' phenomenon.
- To be credible, reliable, and valid, there needs to be reliable measurement at a minimum of two points in time.
- Implementing an assessment plan endorsed by NY State Education Department could actually be a relief to some administrators who have been uncertain about how to approach accountability.
- There needs to be careful watch over the impact of assessments on the classroom. If the measures are too cumbersome and time-consuming, or if they interfere with classroom process, it will pose problems to the acceptability and soundness of the evaluation.

Should the Evaluation Include All Districts?

There were two opinions voiced on the question of whether or not all UPK districts should be required to participate in a statewide evaluation. On the one hand, voluntary district participation was endorsed because full participation may not be technically necessary. If teachers, parents, providers, and policymakers believe that the (voluntary) assessment is relevant, then districts will likely see merit of the statewide evaluation. Offering the opportunity to take part will likely result in a high participation rate.

On the other hand, voluntary district participation could weaken the evaluation because the selection factors are too strong (for example, districts that have poor programs are not likely to choose to participate). This reasoning endorses mandatory participation for all districts (but as agreed earlier, a selected sample of children would participate, not the entire population in a district).

Is There an Appropriate Comparison Group?

Differences across districts suggest that comparison children should be drawn from within districts (as opposed to comparison districts). Finding a ‘pure’ comparison group of children who are not in UPK and are not receiving services from other providers is likely to be impossible.¹⁵

The lack of a true non-treatment comparison group argues for the collection of extensive background information on children and families, as well as information that documents the non-UPK services they receive. Politically, it remains important to tease out the unique effects of UPK over other programs in order to answer the question of what UPK provides beyond what is offered in other programs. Are there aspects of UPK that are unique to the program, such as teacher certification, parent involvement, staff development, or a focus on literacy? These could serve as the basis for identifying data that could differentiate effects across programs.

The half-day nature of UPK compounds the issue of identifying and describing the children’s experiences. Because UPK is funded for 2.5 hours per day, many UPK children are attending other, add-on care. This presents a whole new dimension to the comparisons: children in both the UPK and comparison groups may be receiving not one but several ‘treatments’.¹⁶

¹⁵The continuing existence of New York’s Experimental Prekindergarten (EPK) adds an interesting twist to this situation: comparison children could conceivably be attending this other state-funded preschool model. In districts where UPK and EPK have been blended (as in Buffalo), providers perceive attempts to disaggregate the data across the two programs as a step backwards.

¹⁶Given these issues, one alternative type of evaluation might be to explore the ‘value added’ to programs by UPK dollars. With these funds are they able to boost teacher training, compensation, or otherwise raise quality? Such changes may significantly raise quality but may not be measurable or appreciable.

As participants grappled with the challenges of formulating an evaluation framework, Gary Henry recalled the words of renowned evaluation methodologist Lee Cronbach.¹⁷ He suggested that large-scale, comprehensive evaluations are not necessarily the best approach. Rather than conceptualizing studies as launching the equivalent of a massive ocean liner, researchers should think of sending out a fleet of smaller, more easily controlled studies. Following this metaphor, UPK evaluators might want to take advantage of different but linked evaluative strategies. For example, a performance measurement system with descriptive assessments of children could combine with parent opinion surveys in some districts. In this model, the underlying indicators would still hold, whereas the other data gathered would be coordinated but would not be exactly the same. This approach could allow for more inclusion of qualitative, descriptive analyses to complement the quantitative data.

Next Steps

As the meeting drew to a close, participants, most particularly providers, were urged to look ahead to the next steps. Assuming that the group goes forward with the evaluation framework discussed in this meeting, a consortium and a process need to be identified: Who should be involved, and how? The group raised several points:

- The ‘who’ and the ‘what’ questions are intimately connected. Consensus on key indicators will determine who is involved; by the same token, who is involved is key to establishing the legitimacy of the ‘what’.
- ‘Wave I’ districts should be involved, as they have the most experience with UPK. They should play a significant role in choosing indicators and selecting instruments.
- Representatives from a range of stakeholder groups should be involved as well: researchers, parents, legislative staff, professional organizations, preschool special and bilingual education, kindergarten and first-grade teachers.
- The extensive list of who should be involved suggests the need for a Steering Committee, no larger than the group at this meeting (30 individuals). The Steering Committee could develop proposals, name reviewers to critique the proposals, then disseminate final recommendations to regional groups.
- All recommendations are overshadowed by the lack of funding for an evaluation. UPK Coordinator Cindy Gallagher suggested that determining how much private and local money (from foundations and local school districts) is currently being invested in evaluation could inspire legislators to

¹⁷Cronbach, L. (1982). Designing evaluations of educational and social programs. San Francisco: Jossey-Bass.

pitch in. If an evaluation has a chance of including a comparison group it needs to produce some credible results within the next 18 months. The decision to go forward, then, must be made within the next three to six months.

- A more immediate goal is to circulate the meeting summary draft for feedback in the coming several months, followed by broader dissemination to stimulate interest and activity.

In closing, Mon Cochran cited the challenges ahead. The evaluation of UPK promises to be scientifically complicated and logistically enormous. The group was charged with being realistic about goals, design, and expectations, but also to adhere to the ideals set forth early in the meeting. The meeting was closed with a pledge to keep all those present involved in the process.

Postscript

Following the meeting, several participants commented on the sharp contrast between earlier and later discussions over the course of the meeting. The morning sessions were marked by the insistence that any outcomes-based evaluation be developmentally and culturally appropriate. Then, in the afternoon, there was an apparently easy willingness among providers to accept standardized measures and traditional methods of data collection. This tension between ideals and practical constraints must not be ignored in favor of a “quick and dirty,” “one size fits all” solution. Every effort must be made to maintain a standard for child-outcome assessment that is multi-dimensional, following Love’s model, and is set in an ecological framework that includes issues of family, teacher, school system and community readiness.

List of Participants

<i>Name</i>	<i>Address</i>	<i>Telephone</i>
Elizabeth Boxer	Research Assistant, Cornell Early Childhood Program	607-255-0843
Mon Cochran	Professor, Human Development Director, Cornell Early Childhood Program	607-255-2260 607-277-0612 Home Office mmc6@cornell.edu
Phoebe Cottingham	Smith Richardson Foundation 60 Jesup Road Westport, CT 06880	pcotham@srf.org
Elaine Dentico	Office of Early Childhood 428 City Hall Buffalo, NY 14202	716-851-3626 ie40@aol.com
Marge Dill (Meeting Facilitator)	Director, Tompkins County Human Service Coalition 100 West Seneca Street Ithaca, NY 14850	607-273-8686 mfd3@cornell.edu
Joanne Florino	The Park Foundation, Inc. P.O. Box 550 Ithaca, NY 14851	607-272-9124
Laurel O. Frasier	Early Childhood Admin. Office of Early Childhood Ed. New York City Board of Ed. 131 Livingston Brooklyn, NY 11201	718-935-4255
Cindy Gallagher	Leader, Early Childhood Coordination Team State Education Department Room 381 EBA Albany, NY 12234	518-474-5807 cgallagh@mail.nysed.gov
Denise Gomber	Ithaca City School District Gunderman Road Ithaca, NY 14850	607-277-3043 dgomber@icsd.k12.ny.us

Beverly Gorbert	Ithaca Community Child Care 579 Warren Road Ithaca, NY 14850	607-257-0200
Gary Henry	Andrew Young School of Policy Studies, Georgia State University P.O Box 4039 Atlanta, GA 30302-4039	404-651-2342 gthenry@gsu.edu
Dirk Hightower	Primary Mental Health Project 575 Mt. Hope Avenue Rochester, NY 14620	716-273-5957
Karen Howard	Syracuse City School District 312 Oswego Street Syracuse, NY 13204	315-435-4276 khoward@freeside.scsd.k12. ny.us
Sally Imbimbo	Bank St. College of Education 610 W. 112th Street New York, NY 10025-1891	212-875-4498 simbimbo@bnkst.edu
Kim Kopko	Graduate Research Assistant, Cornell Early Childhood Program	607-255-0843 kak33@cornell.edu
Sarah Lane	Graduate Research Assistant, Cornell Early Childhood Program	607-255-0843 sk124@cornell.edu
Kristi Lekies	Research Associate and Coordinator, Cornell Early Childhood Program	607-255-2457 ksl10@cornell.edu
Diana Levy	Ithaca City School District Gunderman Road Ithaca, NY 14850	607-277-3043 dlevy@icsd.k12.ny.us
Patti Lieberman	Vice Chair A.L. Mailman Family Foundation 707 Westchester Avenue White Plains, NY 10604	914-683-8089 psl101cpw@msn.com
John Love	Mathematica Policy Res. Inc. 600 Alexander Park Princeton, NJ 08540	609-275-2245 jlove@mathematica-mpr.co

Luba Lynch	A.L. Mailman Family Foundation 707 Westchester Avenue White Plains, NY 10604	914-683-8089 llynch@mailman.org
Wilma Martin	Board of Education Ithaca City School District Lake Street Ithaca, NY 14850	wsm1@cornell.edu
Susan Megna	Early Childhood Coordination Team State Education Department Room 381 EBA Albany, NY 12234	518-474-5807 smegna@mail.nysed.gov
Joanne Mulig	Early Childhood Program WIEC 422 Munford Street Schenectady, NY 12307	518-370-8228 muligj@schenectady.k12.ny.us
Anthony Raden	Associate Director Institute for Child and Family Policy, Columbia University 622 West 113 th St. Mail Code 4600 New York, NY 10025	212-854-5448 ar495@columbia.edu
Sue Rakow	Teacher, UPK site 422 Chestnut Street Ithaca, NY 14850	607-272-3882 sfvr@netscape.net
Carol Ripple (recorder, report writer)	Carol Ripple Consulting 25 Dillon Road Woodbridge, CT 06525	203-222-6222 carol.ripple@yale.edu
Judy Ross- Bernstein	Senior Lecturer, Human Development Cornell Early Childhood Program	607-255-0826 jr42@cornell.edu
Grace Schaefer	Office of Early Childhood 428 City Hall Buffalo, NY 14202	716-851-3626

Sally Selden	Maxwell School of City of Public Affairs 306 Eggers Hall Syracuse University Syracuse, NY 13244-1090	315-998-5014 scselden@maxwell.syr.edu
Kathy Skowyra	State Communities Aid Assn. 150 State Street Albany, NY 12207	518-463-1896 ext. 31 kskowyra@scaany.org
Shawn Smith	Ithaca Community Child Care 579 Warren Road Ithaca, NY 14850	607-257-0200 program@icthree.org
Elizabeth Stilwell	Director, The Early Childhood Center at Cornell; Cornell Early Childhood Program, Director of UPK site	607-255-3288 els13@cornell.edu
Fasaha Traylor	Senior Program Officer Foundation for Child Development 145 E. 32 nd Street, 14 th Floor New York, NY 10016	212-213-8337 fasaha@ffcd.org
Christine Vogelsang	Syracuse City School District 312 Oswego Street Syracuse, NY 13204	315-435-4276 FAX 315-435-6553 cvogelsa@freeside.scsd.k12. ny.us
Wendy Wingate	Teacher, The Early Childhood Center at Cornell	wow1@cornell.edu 607-255-3288
Anne Wolf	Division of Assessment & Accountability New York City Board of Ed. 110 Livingston Street Room 740 Brooklyn, NY 11201	718-935-3772 awolf@nycboe.net