

Frontiers in Child Care Research
Annual Meeting of the Child Care Policy Research Consortium
Radisson Lord Baltimore Hotel
March 8–11, 2005

Measurement Assessment and Outcomes
Plenary Panel Summary

Moderator:

Ivelisse Martinez-Beck
Research Coordinator
Child Care Bureau
Administration for Children and Families, DHHS

Panelists:

Martha Zaslow
Vice President, Research
Child Trends, Inc

David Edie
Public Policy Education Specialist
Wisconsin Child Care Research Partnership, University of Wisconsin–Extension
Technical Assistance Specialist
National Child Care Information Center

Donna Bryant
Associate Director and Senior Scientist
Frank Porter Graham Child Development Institute
University of North Carolina at Chapel Hill

Ivelisse Martinez-Beck

The topic of this panel is on measurement, assessment and outcomes, specifically with regard to professional development, quality rating systems and assessment of children.

The measurement of many variables is important to the study of childhood policy. Some advances have been made in recent years, but we are still struggling with how best to measure variables that we know affect child care quality and how well children do in their child care settings. For example, what is the best way to measure stability of employment or stability of

child care use? When we think about measuring stability of care, for example, are we considering the effects of stable arrangements on both parents' employment and children's well-being? How we measure the variables must relate to the questions we need to answer.

What about measures for professional development of the work force? Which strategy will help us identify thresholds of professional development that are linked to quality of care? Which professional development experiences will have an impact on the caregiving practices of caregivers and teachers?

One area needing improvement has to do with how to measure the quality of child care environments *per se*. Quality-of-care measures are increasingly used in the implementation of State-level initiatives such as increased compensation for caregivers who complete certain professional development activities or the designation of special high quality programs.

Many researchers continue to explore the relationship between structural indicators of quality, characteristics of the care environment, and the practices of individual caregivers. Many of us feel that more work is needed to better understand which structural features contribute to the caregiving practices that have been identified by research as supporting children's positive development in various domains. For example, to what degree are learning materials and supplies related to the kinds of teaching practices that have been identified as supporting children's early learning and contributing to specific outcomes? The assessment of programs and children has taken on renewed importance as States and communities are held accountable for demonstrating the impact of their early education programs. In response to requests from States, this topic will be the focus of some of the Child Care Bureau's training and technical assistance efforts within the regions and the States this year.

Some of the questions we hear about the assessment of programs and children include:

- Which features of programs should be assessed?
- Which measures best capture the practices of teachers and caregivers in diverse settings?
- How do these features of programs and care practices link to outcomes for children attending those programs?
- What is the right timing for assessment of children's learning and readiness?
- How do we properly assess children coming from different racial and ethnic and language backgrounds?
- Which measures should be used for which purposes?

All of these questions – and many more like them - have implications not only for children's well-being but also for policy decision making at national, State, and local levels.

Our panelists this morning will be sharing some of the issues raised in previous meetings that continue to be unresolved to a certain degree. They also will present some possible solutions or approaches that have been taken to meet these challenges.

Our first panelist this morning is Martha Zaslow. Marty is Vice President for research at Child Trends in Washington DC. Her work focuses on implications of welfare policies for families and

children, on programs for children in poverty, maternal employment, and child care, and on improving survey measures of parenting and of children's development. She received her PhD in developmental psychology from Harvard University.

Our second panelist is David Edie. Dave has a master's degree in education and is a policy and education specialist at the University of Wisconsin - Extension as part of the Wisconsin Child Care Research Partnership. He has been a lead planner on child care issues in the Wisconsin State government and played a key role in the development of Wisconsin's welfare reform program, Wisconsin Works. Dave held several positions during his tenure with the State, including Director of the Office of Child Care and the Director of Regulation and Licensing. Dave also works as a technical assistance specialist for the National Child Care Information Center.

Our third panelist this morning is Donna Bryant. Donna is Senior Scientist and Associate Director of the Frank Porter Graham Child Development Institute at the University of North Carolina - Chapel Hill. She is also a research professor at the School of Education. Donna is currently leading a five-State team of researchers in a study funded by the Child Care Bureau to assess implementation and outcomes of a quality enhancement intervention for family and center based providers. She led the study team for North Carolina's Smart Start Initiative and has conducted several studies as part of the Head Start Quality Research Consortium. She leads a team that is part of the second Head Start study of mental health interventions for teachers and parents of children with challenging behaviors. Donna has conducted Statewide studies of North Carolina's kindergarten and public pre-school programs, a school readiness assessment pilot study, and a Statewide early childhood needs and resource assessment. She holds a PhD in experimental psychology from the University of North Carolina at Chapel Hill.

Martha Zaslow

Good morning I want to begin by saying that we're viewing the session on defining and measuring professional development today as one stepping stone in a series of stepping stones that have already been laid on the path of understanding what providers need to ensure that children thrive under their care. Others who are here in the room have very much participated in these communal efforts. We hope that the breakout sessions taking place in this meeting will provide the feedback and impetus for sustained progress in these new directions.

The issue of defining and measuring early childhood and professional development has moved from peripheral vision to focal vision in a series of meetings that the Child Care Bureau has sponsored in collaboration with other federal agencies, especially ASPE (Office of the Assistant Secretary for Planning and Evaluation). In particular, I want to share the results of two earlier meetings that put important stepping stones into place.

The first was a meeting in 2003 that dealt with early childhood professional development as a contributor to children's school readiness. This was a meeting sponsored by the SEED consortium of federal agencies with a focus on early childhood. SEED stands for the Science and Ecology of Early Development. The second was a workshop on definition and measurement of professional development and training that took place a little over a year ago. And there was also

a discussion of the importance of definition and measurement of early childhood professional development at the Child Care Bureau Policy Research Consortium meeting two years ago. So we're really talking today about placing stepping stones across the pond and how to make good progress across it.

At the SEED meeting on professional development as a contributor to children's school readiness, Kelly Maxwell, who is here at this meeting, presented a paper along with Cathie Feild and Dick Clifford that was intended to get everyone on the same page about terminology and put some issues about measurement on the table. But instead it put the issue of definition and measurement front and center as a major problem and a major issue. To summarize briefly the paper said that the research is muddled by a lack of agreement on basic definitions of measurement approaches. There is no clear distinction especially between education and training. People use these terms in overlapping ways, and if you try to summarize the research, you're summarizing apples and oranges. Some people refer to training as higher education. Some think of training as workshops - single session workshops, intensive workshops, there is no common terminology. And it makes it very perilous to summarize the research. And training is a particular gap. Research is collected on training but then researchers usually discard it because they find that there are problems with recall and reporting.

When asked about their training experiences, people will remember an enormous range of sessions. They don't have a metric for recording the sessions, and there's no common categorization of content. So people throw away content information. We are lacking a parsimonious description of content of training. We have 30, 40, 50 different categories that people have used in various studies, and there is absolutely no agreement so you can't summarize across studies. There were also problems with how different people define education, which you would think would be very simple and straight forward. For example, people don't distinguish between some college, one course in college, or almost enough hours to graduate. There are also problems with definitions of credentials, especially with people not specifying what's going on in State and local credentialing programs. In all of these instances, we don't know what the content is.

The second meeting, the February 2004 workshop, was intended to build on Kelly Maxwell's paper (with colleagues) along with a new paper by Rick Brandon, Erin Maher and Gretchen Stahr Breunig at the University which extended the picture. Rick and Erin are here today.

The discussion at that workshop made it clear that these problems with measurement have important policy and practice implications. That is to say, we're not just being obsessive or overly particular and fastidious researchers. There are really important applied implications of these issues on the ground. But it also pointed to some promising steps for making progress. I'm just going to quickly summarize some of the key points of the paper by Rick Brandon and Erin Maher and I refer you to their website¹ for copies of the paper.

- There are serious limitations in our efforts to collect data that accurately estimate the size and characteristic of the early child care work force. Flaws in major federal surveys include the fact that they don't include all categories of the early childhood workforce.

¹ http://www.hspsc.org/publications/early_ed/ccw_May_2002.pdf

They tend to eliminate self employed and people who don't self identify as professionals. So family, friend and neighbor caregivers are excluded, as are many other home based caregivers.

- It is difficult to delineate the borders of the work force. For example, we don't know if various specialists should or should not be included. And what about people not directly working with children, such as administrative staff, bus drivers, food workers or custodians?
- At the State level, not all States are collecting data on the work force and there's a lack of comparability across States in what is collected.
- In their review of State surveys, Rick and Erin also found problems not only with measurement but with data collection issues like adequate response rate, how frequently data are updated, and who is included in the data? Does the sample include only people in programs like the subsidy programs or is it a broader survey?

The participants at that meeting pointed to important implications for policy and practice and emphasized that this is an applied issue, not just an example of research fastidiousness. For example, reliable information on numbers in the work force including all sectors is the foundation for efforts to improve professional development and efforts to expand supply. And the lack of agreement on the key dimensions of training is a very important concern because we don't know the content and we are failing to relate it meaningfully to quality issues. For example, Linda Smith, the executive director of NACCRRA, (National Association of Child Care Resource and Referral Agencies) noted a concern that people are taking one training course that they really like over and over again. In some cases, you can fulfill the same hourly licensing requirement by taking the same course again and again.

A better definition of content in training and formal education would help us address key policy issues. One that is on the table now is whether a BA is necessary. But a BA is a marker or a summary variable that we need to un-package and look at. What is the content of the BA in early childhood? What are the nature and the quality of the courses? Another applied issue is what aspect of professional development should be included in quality rating systems? So those are just some of the issues discussed at that meeting.

A number of possibilities also emerged at that meeting that we're following up on now and hope that the breakout sessions this week will be a further step. There's interest among federal agencies that sponsor national surveys and studies in working toward common definitions, improving the categorization and improving the questions. For example, revising the relevant questions on the ECLS-B (Early Childhood Longitudinal Survey – Birth Cohort) is one possibility and there is a precedent in the common core measures used by the NCES (National Center for Education Statistics) for K-12. We could follow their example and develop a common core of measures used in national surveys and in State surveys. And there is growing interest in this approach.

The Child Care Bureau's Quality Interventions for Early Care and Education (QUINCE) project is field testing new measures for provider training and linking training variations with child outcomes. What we're hearing from the field is that we can go too far with these measures. So

there needs to be a triangulation of what is real and important, what can really be collected, and what will also enable us to make progress.

There's also the possibility of further analyses with existing data sets. At the break out group you'll be hearing from Linda Smith of NACCRRRA that they are first of all collecting a survey of their directors on training. And there's also a possibility of including the same variables in NACCRRRAware, their new national software information that can categorize training content at individual provider levels.

The National Organization of Child Care Registries is also represented here today. Jere Wallden and Kathy Thornburg will be at the next break out session to talk about an initiative that has to do with using common content areas in registries across States. So this is also very exciting.

The Child Care Bureau's State data and research capacity-building grantees, with Bobbie Weber and others, who are here today, are working on a common identifier to link data sets that have important information about the work force in order to get a more comprehensive and complete picture of the work force across States. And there is the potential to improve data collection in market rate surveys which must be collected on a regular basis. Perhaps these could be expanded in a meaningful way to include data on the characteristic about the size of the work force.

After these early meetings, a workgroup on definitions and measurement of professional development was formed to follow up on and move forward in resolving these issues. It was formed, it does exist now, and it is an open, active forum. We welcome participation from those who are particularly interested in this issue. Our intent is to maintain open communication so that, to the extent possible, these new efforts will inform each other and promote steady progress. A key principle underlying the work of this group is that of encompassing multiple sectors of early care and education. This is not to be just child care. We want special education and child care and Head Start and pre-K to be communicating on these issues so that we move forward in a collaborative way to develop a common frame work for identifying key dimensions and measures. We also have a goal of learning from initial efforts by testing promising approaches in the field, and we have to do an honest assessment of pitfalls encountered in field tests and in reality.

One of our worries is that there might be a wonderful research measure but when you put it out in the field, and there are high stakes associated with it, you may actually get distortions of the data. That is to say, measures can work in their proper context but may get distorted either by being used improperly or through careless interpretation, so that it's not just the precision of the measure but also how it's used that's important. Sue Hegland from Iowa State University is working on a paper that examines these issues.

So very briefly now here is the frame work for the way we want to think about professional development and training. We have further thoughts about how to do this and would welcome your input. We want to look at the content - to go beyond the BA degree *per se* and talk about the nature of the course, the substance of the course. Are courses in literacy related to literacy outcomes in children or the literacy environment? Let's go beyond the imprecise measures that we're using.

We want better measures of dosage or extent - for example, the single workshop vs. the intensive series of workshops. We want to look at modality. What do we get with distance learning as opposed to in-person training? This is very important in rural areas. How are our efforts at distance learning working? We need to know whether professional development is more effective when linked with a system like pre-service requirements or whether it is something that the provider or the educator can do on her own? This may make a difference. The motivation and the context may be different.

And finally we think it's terribly important to think about requirements to demonstrate competence. Are we thinking only about learning and information or also about practice?

The presenters at the next break out session will be speaking today about follow up efforts that are now moving forward. Linda Smith from NACCRRA will be talking about efforts to document training through a survey of directors of resource and referral agencies and about integrating detailed measures of professional development content into NACCRRAware. Jere Walden will be here to talk about efforts of State child care registries. Kathy Thornburg will be there as well, and Kathy has been very involved in coordinating definitions used in registries. Bobbie Weber will be talking about the use of a common identifier to strengthen State data collection on work force using administrative data. Donna Bryant will be talking about extending measures of content, extent and purpose of professional development in the QUINCE evaluation - including the reality that what a field test tells you might be that you are being overly ambitious. And finally Rick Brandon will be talking about next steps in State and federal surveys. Your input is very much requested. We need to know how these issues are emerging in your State, your organization and your research. Do you see particular problems that we haven't noted here? How are these issues affecting policy and practice from where you sit? And do you see further opportunities for progress?

David Edie

I'm going to be focusing my remarks on research and the road to policy. Just as an aside, I'll tell you that a couple of decades ago my brother worked in the US Senate for many years as an aide. I called him and asked how often research is crucial to a political decision in the US Senate. And he said never. Now, I'm hopeful that we've reached a new world in early care and education where research is extremely important. So I'm going to talk about the connection between research and policy.

My background is this: I am not a formal researcher but in my two child care roles I have gained a sense of how research relates to policy issues. First, I've been working with the Wisconsin Child Care Research Partnership helping the State develop a proposal for quality rating systems. That position has offered me an inside view of a State's efforts to develop a quality rating system. And secondly I work for the National Child Care Information Center where we've been collecting information about quality rating systems from across the country, so we have a lot of information across the States. This morning, I'm going to look at this from the point of view of a State administrator thinking about a potential quality rating system. And as you know, this is a very hot topic right now. In my presentation I'm going to talk about four questions that I think

are key to what States are facing, that I think should be of interest in framing the issue: the why, where, how, and what questions.

Why should a State create a quality rating system? Every State has its own reasons, so I want to emphasize that this is a very State-specific venture. In everything you do you need to be aware of where your State is on this issue. First of all, we know now from scientific evidence about the importance of early years in brain development. Of course, research just reinforced what we already knew from past research and life experience. School readiness and achievement gaps are a major driving force of many States. There are growing concerns that quality of child care and early education are not what they should be or could be. I believe that all children deserve high quality and evidence that early investment pays off.

So where are you now? This is the assessment question that States need to ask and are asking. Where are you as a State and what do you need to know before you enter into a quality rating system? I think this is crucial. Policymakers in almost every State that is working on a quality rating system are trying to do an assessment of where they are and what information they have. What research do they have that speaks specifically to their State, or do they think national research is relevant and can be used?

What's the level of quality that you're starting at in your State and what are you trying to build on? What quality supports are already available and are they systematic? What administrative data are available? We think that these, at least from our experience, are very crucial items, because you need sufficient data to start building a Statewide system and most States are saying that you must automate ongoing data about your quality rating system or you're going to be in big trouble. So this is an extremely important question.

The next question is **how would you do it?** A couple of the key questions are what indicators do you use and how do you measure them? The NCCIC could give you a lot more indicators, but I've picked the top five that they've looked at from a sampling of States and communities with three or more tiers in a quality rating system. These top five indicators that States are using include: (1) educational qualifications, (2) learning environment and curriculum, (3) parent and family involvement, (4) accreditation, and (5) licensing compliance and status. But there are many more indicators that States are looking at and there are lots of different ways to measure them.

As our team talked about this presentation, we thought about the unusual connection between research and broad public policy. Research tends to be cautious, nuanced and precise. You know researchers are always debating terms and definitions and everything else. Researchers tend to be complex and they try to show the limits on findings. Broad political decisions tend to be much more global and value-based. Unlike research which tends to be expressed with caveats and limitations, policy tends to be assertive, certain, definite, and expressed by slogan.

So we're moving often from precise measures with lots of caveats to a blunt instrument of public policy that tends to be somewhat certain. And very often, that's a very uncomfortable fit. So we've been thinking about a continuum of how to assess quality that takes into account these two ways of approaching the question. What kind of criteria would be used and how would they be

measured along a continuum? I think I've heard threads of this discussion going on quite a bit already at this conference.

At one end it's a simple, inexpensive approach, using proxy indicators. So the question that's coming up in almost every State is whether to use structural indicators of quality that are easily measured and are available through administrative data or other reasonable methods. This is a popular approach because the research that has been done on proxy indicators seems to show that they do in fact reflect differences in quality. On the other end of the continuum are detailed on-site assessments which tend to be complex and costly. So the continuum moves from simple indicators to in-depth measurement of the interactions that structural indicators are intended to be proxies for. And somewhere in the middle are approaches that use a combination of proxy indicators and more sophisticated measures-- maybe some limited observations. And of course this is a very, very difficult question. We're not saying that we have an answer, but we think it's going to be a very hot topic in States as they're trying to control costs and be accurate.

Finally we wanted to ask what **criteria you would use in building a quality rating system**. What are the values you're bringing to it? What are the criteria? In Wisconsin we developed our own criteria, and again I think every State is different but we seem to fit the situation in Wisconsin.

- We wanted the system and the indicators to be simple, clear, understandable, and easy to use, particularly for parents. We wanted the quality indicators to be valid, fair, objective and valid. We wanted to be able to make good policy decisions based on data from these indicators. Those words flip off the tongue but doing this well is a very big chore.
- We wanted the system to be realistic and structured so that programs could reasonably move from one step to another up tiers of quality, and we wanted the system to be efficient, of low cost, and easy to administer. If you play with these concepts enough you realize that there's a lot of conflict and tension between all of them.

So this was the vision. We're hoping that through sound research findings, good data, and a planned, analytical approach we can arrive at public policy that has systematic positive impacts on young children in their early care and education settings. That's our framework for thinking about quality rating systems.

If you would like more information about how we did it, you can download a copy of our report entitled *Developing a Child Care Quality Rating System: Wisconsin's Approach* at www.uwex.edu/ces/flp/wccrp/publications.html

Donna Bryant

I want to mention four main topics in my remarks: The purposes for which children are assessed, the timing of assessments, the assessment procedures (which obviously covers a huge range of sub-issues including who does it, which children, when, how, and which measures), and lastly, some questions I have about costs and benefits.

(1) Purposes for Assessing Children

In terms of the purposes for assessing children, as a field I think we've done a very good job over the past decade since the National Education Goals Panel issued its recommendations. Several documents have been produced that delineate the purposes of a variety of assessments, and their main points are almost like mantras that we're all able to say: the most important aspect of an assessment system is the linkage between the curriculum and children's skills and knowledge. Or if we're talking about evaluation we make it very clear that assessment is not the same thing as evaluation. I think that we've done a good job in delineating this but some people just don't listen, and so we're still put in the position of often being asked to use measures for purposes for which they were not designed.

A case in point: we're part of a well-funded study sponsored by an unnamed organization. All of our sites are using some kind of screening instrument with the children who are part of the study, and we were asked at one evaluators' meeting to provide standard scores. This request was adamant – almost to the point of demand. But we evaluators were firm in our answer that we could not provide standard scores. We can talk about screening. We can talk about the proportion of children that are at risk. We can talk about the proportion of children that look okay. But we're not going to use these measures that were intended for screening purposes as a high stakes outcome evaluation. I have to say that was very, very hard to do, and we were only able to do it because there were lots of us in the room. But I think it's an issue for all of us. Even though we've delineated it, the message isn't always being heard or accepted.

(2) Timing of Child Assessments

The second issue I want to address is the timing of when child assessments are conducted. And here I'm particularly thinking about the use of assessments in evaluations. I'll give Smart Start in North Carolina as an example. Two members of that evaluation team, Kelly Maxwell and Karen Taylor, are in the room today. As we worked throughout the first years of Smart Start, one of our biggest successes in my view was actually delaying any kind of child assessment because all of the efforts were going into improving services, improving quality, getting more slots for child care, getting more children in high quality child care, increasing professional development. Those are wonderful things to do, but if you're going to see a broad change in children you have to go for quite a while or with great, great intensity, neither of which was happening in 1994, since the program just started in 1993. So we kept saying to our State policymakers, no, no, no let us show you changes in these immediate outcomes and then we'll assess child outcomes. I think that's another case where we won by pushing back and saying "don't shoot yourself in the foot by assessing children before they're ready, before your program is ready to be evaluated in that way."

(3) Procedural Issues in Assessment

The third point I want to briefly make, and that will be talked about in much more detail in the breakout this afternoon, is that there are many, many procedural issues about assessments that challenge all of our evaluation designs. Which children shall we assess? Which measures shall we use? Who is going to administer the assessment? Often we must face the issue of whether to

asses all children or a sample of children. This issue usually occurs during the design of program evaluations where we know very well that a sample will do. In the Statewide North Carolina School Readiness Assessment, the power analyses surprised me: In our tenth largest State in the country, a well-constructed sample of five hundred children gave us a very good estimate of children's abilities. We actually doubled the sample to a thousand just because a four-digit number seemed more robust than a three digit number, but the truth is it didn't buy us much more power. So it's really hard, I know, to convince policymakers that sampling will give you perfectly adequate answers in many cases. It certainly won't help you guide an individual child's curriculum and activities in a classroom, but for evaluation purposes a sample will suffice. That was another victory. Governor Hunt was a good listener. He had initially said, let's do one measure with all children, but we convinced him to do multiple measures or sub-measures with some children. That was an important assessment victory.

Another procedural issue concerning assessment is the age at which children should be assessed. Those of us who are dealing with children across the age range from birth-to-five know there is no perfect measure that will let us watch children's development from one age to the next. In the QUINCE study we had chosen the PLS (Preschool Language Scale) because it is standardized from birth to six. But the individual infant/toddler items are really quite different from the items for four- and five- year olds. So even though measures may be called by the same name, I'm not really sure that they're giving us what we want when used with younger children.

Children who are limited in English proficiency are another group of children for whom we have much improved practices. We've developed really good procedures in our research and evaluation studies for sort of triaging children. Who should get what kind of assessments? For example, I think we do a very good job of appropriately moving children into a Spanish-based assessment or an English-based assessment. Of course, we don't go much beyond that and we also end up with two separate sets of measures. What do we then do with the data? How do we interpret it? We've at least assessed the children more appropriately than we used to, but we know that children who are learning two languages or switching from one language to another are very likely to be depressed in both languages on any kind of assessment you use. So I think we're administering better, but I don't think we're necessarily using the data we get as well as we will need to do in the future.

I think that the indicators that the National School Readiness Indicators Consortium of 17 States has been working on should be high on our list of potential outcome measures. Anytime you can use indicators from large databases it lends credibility, and indicator data are becoming more widely accepted. The new report, out last month, is at www.gettingready.org

The National Education Goals Panel did a great job delineating the five domains in which we should assess children. However, it is much easier to assess children in literacy and cognitive development than it is to assess their approaches to learning and their social/emotional development; so not only are we getting pressure that tips us toward the literacy and cognitive domain, our own lack of assessment tools tips us in that direction. I suggest that we keep our eyes on the work of Cybele Raver, Lisa McCabe, Jeanne Brookes Gunn and others who are researching ways to measure approaches to learning. They may soon have some measures that can help us.

In my view, we are doing a somewhat better job now, compared to the past, in assessing schools' readiness for children. The dual direction of assessments is important. It's not just how ready children are for school; it's what schools and communities are doing to be ready for them. Many evaluations now include measurement of schools activities vis-à-vis the transitions of children from families into schools and also the continuity between the child care community and the school community.

The final procedural issue I will note concerns who should make the assessments. As a researcher, I like to have independent, well-trained data collectors who are blind to any treatment conditions, but we have lots and lots of studies now where teachers, or mental health coordinators, or people in the intervention or early education systems are collecting data that is being used for high stakes purposes. For determining classroom instruction for individual children, I think such reporters are perfectly adequate and, in many cases, the best people to conduct such assessments. However, when you put high stakes on the assessments, when you ask a teacher to assess a child in the fall and again in the spring knowing it's the gain scores that are going to be looked at to evaluate that program, it's pretty easy to figure out your children should score low in the fall and make a lot of improvement by the spring. So we need to be very, very cautious and continue to talk about the problems of doing these kinds of assessments for those high-stakes purposes.

(4) Benefits and Costs

The last topic I wanted to mention is the issue of the benefits and costs of assessment. This is not my field and I'm going to find Rick Brandon and ask for help, but it seems important to know who's benefiting from the assessments we do with children. Do the children get a better intervention as a result of our assessments? Do they get more services as a result of our assessments? Are the good programs that are serving them now likely to continue if an evaluation using child assessments doesn't show change?

We ought to pay more attention to the cost of obtaining those results. There's not only a large monetary cost in just buying all the materials and arranging all the places and training all the people to conduct assessments, but there are lost opportunity costs. What else could those people be doing with their time in providing direct services for children, or arranging better care for children, or providing better experiences for children, when you have large scale assessment programs for two- and three- and four-year old children? Maybe we need to start directly calculating those costs in our studies. There may be a benefit - a big benefit, a small benefit - of having assessments on every child or on some children. But what are the costs to us both monetarily and in lost opportunities of doing those assessments?

Those are the four issues that I wanted to note in my ten minutes. I refer you to the session at 2:30 this afternoon which will go into much, much more detail and hopefully give us some new directions that we can follow in our research.

Comments from the Audience

Audience Member: I want to ask a question about research on quality and the impact on children. Why do we need to do this? I am a health and policy researcher and I study home ownership. Something very clear to me in my work is that ideologically it's automatically assumed that home ownership is good for people, particularly poor people. The research that's been done - I won't say it's lousy but it's certainly not definitive, and there's a whole lot less research on low-income ownership in terms of its impact on families and children than there is on the impact of quality child care on children. But it's accepted uniformly within many policy circles that home ownership is good for families, it's good for children, it's good for neighborhoods, and it's particularly good for poor people. This policy stance is being driven by ideology not by facts of research. Here we have a child care situation, a situation where it shows that quality is good for children, but yet we're constantly at this point of asking what's the impact of quality on children and doing all these intensive measures. So I raise the question - and I don't expect anyone to answer it - but what is it ideologically about child care that makes it impossible for people to accept the idea that quality may be good for kids without all this measurement? In Congress we have all these other issues where ideology can drive it and there's no problem.

David Edie: A few years ago when I when in France on a delegation looking at the French system, we kept asking them about the research that showed that in fact they had made a difference. And they looked at us as if we were from another planet: This is good for our children we all accept it, what else do we need? So you know you don't have to have all this research if you believe that this is the right thing to do. But we live in a different kind of society that demands results and evidence. So it just may be a different view. But I wonder if we couldn't make the case, because there's so much research linking structural features of quality with eventual good outcomes. Do we have to keep doing this everywhere we go, or can we accept that as a reasonable scientific finding and work on improving the quality?

Donna Bryant: I think to a certain extent the research is turning a corner from beating that dead horse to asking a different question, which is what are the initiatives to improve quality that are worth investing in? I think that's a very important transition and I think it's very exciting to see child care and school readiness evaluations going forward from the SEED consortium and the QUINCE evaluation and the curriculum evaluation. I think in five or ten years we're going to have a whole new set of random assigned experiments emerging from the subsidy evaluation experiments, including the literacy evaluation in Florida and the family child care evaluation. There will be a new body of evidence to summarize and it will point to the next generation of questions about what are the interventions that work and then hopefully how much do they cost and how much do they tell us.

Audience Member: Having just emerged bloodied from a battle to put a quality rating system together, what I'm concerned about is that we have research that identifies a package of assets that tends to be associated with higher quality, but once you take a measurement and you use it for high stakes decisions, as Donna pointed out, it can lose its predictive validity. Here's what I'm seeing happen in the quality rating system battle. To reduce the cost, we drop the structural

characteristics and the process characteristics that are high in cost - like on-site observations. Although training programs may have outcome assessments, we still keep the cheap indicators. So we keep the number of hours of training but drop the outcomes assessment. We keep group size but we lose the accountability. We hire a director with a bachelor's degree over a more qualified CDA or AA degree. We pay her eight dollars an hour. Why is she working in child care? She got a C in Student Teaching and couldn't get a job in the public school down the road. This is a high stakes scenario because in setting reimbursement rates we take the cheap structural characteristics, we drop out the measures of accountability, and we lose the predictive validity of those same structural characteristics.

What I'm arguing is (1) we need to keep doing research to determine whether the assessments still have the same predictive validity that they had before being used in high stakes situations; and (2) we have to include accountability some place in quality rating systems, either in the way we measure the training or in the way we measure the final quality of the process. That means you need to put money into the system. You can't have this system operate (as my State is proposing) with just one registry person who is going to be paid fifteen dollars an hour to enter data and that's what's going to drive the entire quality rating system.

Dave Edie: Well, you're really talking about the continuum I was describing. You go from inexpensive to expensive and you go from structural indicators to complex observation and data collection. Right now, States are going in lots of different directions. Hopefully, if they go with the structural indicator approach, there will be an evaluation to see if it actually results in higher quality. We need to see if those simple indicators actually tie to higher quality as measured by more complex research measures.

Donna Bryant: I think we have to realize that measurement in itself is an intervention and what we measure is what we tend to get. Ann Witte presented results at the last Child Care Policy Research Consortium meeting showing that when licensing results were posted on the web site in Florida, the number of inspectors went up, the number of visits went up, the ratings became more differentiated, and the quality improved. It's very interesting. The quality doesn't seem to have improved because parents walked away from the places with problems; it seems to have improved because the centers knew that their reputations were on the line and what we measured changed their behavior. So we have to be conscious that our measurements do not exist in a vacuum. They exist in practice as they are implemented and the very fact of putting them into place affects practice out there. Measurement can distort what we are measuring, but if we choose the right measures they can also have positive implications.

Dave Edie: I wanted to mention something that I forgot. I wanted to do a promo for a couple of workshops that deal with the issues that I was talking about. Breakout Ten will look at quality rating systems in which several States are involved. Session 11 will look at administrative data which my presentation indicated might be key to successful quality rating systems.

Audience Member: It seems to me that a lot of the comments that we're hearing in this discussion really point up the underlying intersection between goals for children in programs and

policies and the underlying attitudes in our society about the roles of parents at home and in the work place. How parents are thinking about what they want for their kids - which we really haven't heard mentioned here today. And how those measurements can either coordinate with or coincide with or be in tension with the kinds of things that we're measuring at the program and child level in quality. It wasn't so long ago that the big question in child care was whether moms should be working at all. It was right after world war two in the early sixties and I think it's wonderful that in our generation we don't ask that question anymore. But some of the issues about what is acceptable to people in policy and what is acceptable to the public still seem to reflect some of those underlying attitudes.

I just wanted to remind us that a few years ago, the Oregon Child Care Research Partnership showed that parents and professionals define quality in very similar ways, and that one of the goals of this meeting is to explore the intersection and the cross-cutting nature of our themes, not only across programs but across populations and constituencies. So I just wanted to raise that as an issue for measurement and for understanding why we have such a hard time getting our work into the policy realm and into public acceptance.

Ivelisse Martinez-Beck

If there are no more questions, I want to remind you about our sessions the rest of today and tomorrow.