Roundtable on Measuring Quality in Early Childhood and School-Age Settings: At the Junction of Research, Policy and Practice

December 4-6, 2006
Washington, DC

Meeting Summary
Draft of July 5, 2007

This document was prepared under Contract # HHSP233200500198U with the Administration for Children and Families, U S Department of Health and Human Services.

Prepared for:
Ivelisse Martinez-Beck, Ph.D.
U.S. Dept. of Health and Human Services
Administration for Children and Families
Office of Planning, Research and Evaluation
370 L'Enfant Plaza Promenade, SW
Washington, DC 20447

Prepared by:
Child Trends
4301 Connecticut Avenue, NW, Suite 350
Washington, DC 20008

Executive Summary

Purpose of the Roundtable

Quality measures were originally developed for research aimed at describing the settings that children spend time in and identifying the characteristics of these environments that contribute to children's development. Increasingly, however, measures of quality are being used for further purposes. In particular, they are being used to guide improvements in practice and as components of state policies.

Regarding practice, in many initiatives to improve the quality of early and school-age care and education, a starting point is the completion of an observational measure of the quality of the environment. Results are then used to identify specific areas in need of strengthening and to develop a plan for quality improvement.

Regarding policy, states are dedicating funding to improve the quality of early and school-age care and education, often going beyond the required four percent set aside of Child Care and Development Fund to allocate further state and federal funds to this goal. One increasingly prevalent approach across states is the development of quality rating systems. These systems provide summary information (for example, with five stars indicating the highest overall quality) regarding the quality of all licensed early and school-age care and education settings, or a subset that chooses to participate in a voluntary system. The intent is to increase consumer information and awareness, and to improve quality through the provision of technical assistance and resources, and through increased demand for quality settings.

State quality rating systems reflect a state's consensus on defining and measuring quality. Most state quality rating systems include the direct observation of quality as a major component of the summary rating of quality. They often include other measures of quality as well, including such structural measures as educator/caregiver education and training and group size or ratio.

The widespread use of quality measures in practice and policy as well as in research is raising a new set of questions and challenges for the early and school-age care and education field. A roundtable was convened in December, 2006, in Washington, DC to bring together representatives of the research, policy and practice communities to engage in a dialogue about issues that are emerging with the increasing use of quality measures for differing purposes.

Key Questions Addressed at the Roundtable

- Using observational measures of quality to guide improvements in practice rests on the assumption that the measures are capturing the most important facets of quality for children's experiences and their development. What does the research say about the strength of our existing measures of quality? Are these measures zeroing in on the aspects of quality that need to be emphasized if we are

interested in strengthening children's school readiness and progress in school? Are we missing any important facets of quality in our current measurement approaches?

- Is there good agreement across different states in terms of what measures of quality should be included in state quality rating systems? Do states agree on what is "good" quality?

- Are states willing and able to update their quality initiatives when and if new information about the measurement of quality emerges from research?

- Are states encountering challenges in the application of quality measurement when the measures are used for policy and practice purposes? How are states deciding on such issues as how often quality needs to be observed and in how many classrooms in order for a quality rating to be valid?

- What protections are needed to preserve the integrity of the rating process from bias?

- What steps are states taking to help assure that low-income families have access to early and school-age care with higher quality ratings?

- What approaches are states using to evaluate whether and to what extent the measurement of quality for practice and policy purposes actually results in improved quality, both within particular groups or classrooms, and in the number of high-quality settings within a state?

Themes That Emerged During the Roundtable

The components of quality

- Each state that is formulating a quality rating system must go through a consensus building process with key stakeholders to reach agreement on how quality will be defined and measured within the system. With 14 states now having put statewide quality rating systems in place, and at least 20 more states developing them, it is interesting to note that there are both similarities and differences in the components states have chosen for their systems and how one level of quality is distinguished from another. For example, most states include direct observation of quality in their rating systems, but some states rely entirely on self-reported measures (for example of group size and caregiver education and training).

- Research on the measurement of quality is moving forward and reaching some important new conclusions even as states are moving forward with their quality improvement initiatives and ratings systems to increase consumer awareness of quality. State representatives present at the Roundtable were clear that they are open to new information that can be used to adapt and update their quality initiatives and rating systems.

- As one example of emerging research, findings are pointing increasingly to the need to go beyond general or global measures of quality to include also components focusing on instructional practice, such as supports for early language and literacy development. These extensions of our well-established measures of quality appear to be stronger predictors of academic adjustment and progress than global measures of quality. State representatives at the roundtable were open to the concept of including newer, more in-depth measures of quality related to instructional practices, as well as more global measures of quality, in their rating systems and to guide practice.

- State representatives were also clear that health practices provide a critical foundation for quality (on which global measures and the newer more focused measures targeting instructional practices build). There was consensus at the meeting that quality rating systems should include a specific focus on health to help assure that this fundamental component of quality is rated and reported on to consumers. In some states, the licensing system suffices to monitor this aspect of quality, but there were concerns that in others the quality rating system provides a needed opportunity to complement more limited monitoring and reporting on health issues.

Implementation Issues

- A body of knowledge is emerging within states about the application of quality measures for policy and practice purposes. For example, states have looked at such questions as whether the same summary ratings of quality are arrived at when settings are observed more or less often and when more or fewer classrooms are observed. They are also looking at whether the same person can provide the summary rating of quality for the quality rating system and for the provision of technical assistance, or if these observations need to be kept separate. It will be important to share this important information on the implementation of quality measurement on a broad scale. Participants expressed interest in a consortium that could share new research and information on implementation issues.

- At the same time, practices are emerging across states that pertain to assuring the integrity of rating systems. For example, states are developing practices to assure high levels of reliability in those who observe quality as part of the ratings system, they are carrying out periodic checks of self-report information that contributes to the ratings, and they are putting in place formal procedures for early and school-age care facilities to challenge a rating. Participants at the meeting discussed the potential of documenting not only the components of quality rating systems, but the practices that are needed to assure their fairness.

Effects of quality initiatives and widespread measurement of quality

- We do not yet know whether quality rating systems and other quality improvement initiatives can and do result in system-wide improvement in quality. Participants in the meeting agreed that it will be very important to design and

carry out research focusing on this critical issue. We need to understand whether there are key differences in the approaches states and communities are taking. For example, do we see an increase in the proportion of higher quality care settings when initiatives involve the investment of substantial resources not only in completing and publicizing quality ratings, but also in providing guidance and help to care settings to improve quality? Do we see differences in effects when participation in state quality rating systems is mandatory as opposed to voluntary? Coordinated research across multiple states was recommended, addressing the effects of widespread measurement and reporting of quality.

- Participants at the meeting discussed the key issue of whether quality rating systems would have the effect of increasing access to high-quality care in low-income communities. Some concern was expressed that the price of high-quality care might increase with increased demand once a quality rating system was put in place, making high-quality care less rather than more accessible to low-income families. Some states have put in place incentives or requirements to help assure participation in quality rating systems by those providing care in low-income communities. While the overall effects of quality rating systems on demand need to be studied, a specific focus on low-income communities will be critical.

- We have a body of knowledge on what types of care different groups of families tend to utilize, but much less explicit focus on how the choice of a particular care setting is made. A critical opportunity now exists to study whether and how parents make use of summary ratings of quality in choosing care, and the extent to which there are constraints on their use of such information.

- Participants discussed the importance of understanding how quality initiatives are affecting child outcomes. Some state representatives raised concerns about the pressure to document improvement in child outcomes in order to receive continued funding for their quality initiatives. Participants discussed the challenges of conducting statewide assessments of child outcomes including, for example, the expense of sending trained assessors to collect data from children across the state.

Follow-Up Steps

The discussions at the Roundtable generated a number of issues and questions that will be addressed through new research, literature reviews, research briefs, and a second Roundtable supported by OPRE. Specifically, OPRE is supporting:

- A series of papers that will review different components of quality and its measurement. These papers will address the following topics:
    - The dimensions of quality that have and have not been captured adequately in the measurement of quality and in the research literature

    - The functioning of quality measures designed to assess both center-based and home-based settings
    - The role of quality measures in policy and practice contexts
    - The status of culture in the measurement of quality
    - The measurement of health and safety as a component of quality
    - The measurement of quality in settings serving children across the age range, including infants and toddlers and school-age children
- A Research-to-Practice Brief series based on the literature reviews described above
- A compendium of quality measures with descriptions that facilitate comparison of constructs that are covered and psychometric properties of different measures
- Commissioned analyses of existing datasets to identify the components of quality that are the strongest predictors of child outcomes
- A second Roundtable to go into greater depth on the issues identified in the literature reviews and new analyses of quality and child outcomes

Full Meeting Summary

MONDAY, DECEMBER 4, 2006

Welcome and Opening Remarks

Moniquin Huggins, Acting Associate Director, Child Care Bureau, opened the Roundtable with welcoming comments. She welcomed participants to the meeting on behalf of the Child Care Bureau. She noted that the focus of the meeting will be on the measurement of quality. She underscored the emphasis of the meeting on the translation of research into policy and practice, and the need to identify gaps in the research. She went on to provide an update on the reorganization within the U.S. Department of Health and Human Services, which involves placing the Child Care Bureau under the Office of Family Assistance. There will be continuity in the programs for low income families. The reorganization will tighten the Child Care Bureau's connections with TANF and families moving off of welfare. There will be continued focus on the quality of child care, including for family child care and unregulated providers. The research team will be moving to the Office of Planning Research and Evaluation (OPRE) in the next couple of weeks. The Child Care Bureau and OPRE are committed to maintaining a close connection between the research and program sides of the Child Care Bureau's work.

Ivelisse Martinez-Beck, Research Coordinator for the Child Care Bureau, added to the welcoming remarks from the Child Care Bureau. She noted that our aim will be to cover all of the issues that are critical to the measurement of quality in policy as well as program contexts. This will be a working meeting, with substantial opportunity to exchange information and ideas in order to identify the key issues and next steps. She thanked all of those who made the meeting possible, including Andrew Williams, the Child Care Bureau Research Team, the Quality Roundtable Planning Committee, federal partners ASPE, NICHD, OSEP, IES, and OPRE, colleagues from the National Child Care Information Center and Research Connections, and the staff at Child Trends. She invited participants to introduce themselves briefly.

Looking at Quality through New Lenses: Goals and Process for the Meeting and Overview of Key Issues

Kathryn Tout gave an overview of the goals and processes for the meeting. She noted that the measurement of quality in early care and education is no longer being viewed only through the lens of research. It has also become central in the policy arena, as virtually all states have put into place initiatives to improve the quality of child care. It has become central to practitioners working to meet guidelines or standards for quality in these initiatives.

Tout noted that research indicates quality of early care and education influences children's development, though the influence of parents and families is greater. While many families rely on nonparental care and education from birth through age 12, as

parents pursue employment or seek to enrich their children's experiences, there is concern that much of the care available does not promote optimal child outcomes.

Tout indicated that as quality has moved into the policy arena, many issues and questions have arisen:

- Researchers have been asked how to define and measure quality across different types of early care and education settings, as well as for different age groups.
- Policymakers have been asked how best to allocate resources to improve quality: what incentives and supports make the biggest contributions, and how much should be allocated overall.
- Practitioners have been asked what works to improve quality.

In this new context, researchers, policymakers and practitioners have needed to learn each others' terminology, as well as how to work together.

Within this framework, the goal of this meeting is to support researchers, policymakers and practitioners in learning from each other, identify issues that are emerging, and plan for next steps. The structure for the meeting relies heavily on discussion within small groups. Plenary panels will introduce each new section of the meeting, identifying key issues and themes. Facilitators will then lead breakout discussions designed to go into more depth on specific issues. At the end of each set of breakout sessions, reports back to the full group will be provided so that everyone benefits from in-depth discussions. Each breakout session will include discussants to help spark the discussion, and scribes to record key points from the discussion.

Marty Zaslow and Kathryn Tout then jointly gave a presentation to identify some of the key issues emerging as quality comes to be a focus of policy as well as practice and research. Their presentation, *Measures of Quality at the Intersection of Research, Practice and Policy*, noted that states are clearly making investments in the improvement of child care quality, with many states exceeding the required 4% set aside of Child Care and Development Funds. Quality improvement funding is going towards many different types of initiatives, including those focusing on improving health and safety, creating or enhancing systems of professional development, and establishing and implementing quality rating systems (QRS).

The presentation used QRS to illustrate the kinds of issues that can emerge with the widespread use of measures of quality through a policy initiative. QRS were selected for illustrative purposes since 14 states have now implemented statewide systems, and 24 states are in the process of designing or exploring the possibility of developing QRS.

In the presentation, Zaslow and Tout noted that the widespread measurement of quality through a state or community QRS holds new potential but also poses new challenges. We note here key selected points from the presentation.

The potential contributions of widespread measurement of quality in QRS include the following:

- Drawing together a range of different stakeholders to reach agreement on the components of quality that will be recognized and how these will be weighted and measured.
- Providing consumers with more and better information about quality to use in the selection of care settings.
- Increasing provider awareness of the components of quality and how levels of quality are distinguished; information that can then be used for goal setting and quality improvement.
- Providing a common framework that applies across different types of early care and education programs, including center care, Head Start, pre-kindergarten and home-based care.
- Providing information that can be used for accountability purposes, which in turn may increase the likelihood of public and private investments in quality.
- Providing a framework for diagnosing the greatest needs for quality improvement, and for gauging progress.

The challenges of widespread measurement of quality in QRS include the following:

- Moving forward with implementation of a system before all of the needed research on the measurement of quality is completed.
- The expense and precision required in developing and maintaining a system involving the ongoing monitoring of quality.
- Guarding against the possibility that the ongoing measurement of quality may alienate some providers from the regulated system.
- Assuring that quality improvements occur in neighborhoods with low income families, and that children from low income families have access to the highest levels of quality.
- Assuring the exchange of research and practice information across states and communities that are moving forward with the measurement of quality in the context of QRS.

Zaslow and Tout noted that key variations are emerging across different state QRS. For example:

- Some state QRS set criteria for levels of quality that must *all* be met before a provider can move to the next quality level, while other state systems involve a tally of points across quality criteria, with quality level established in light of total number of points rather than the need to meet a specific set of criteria;
- Most states tie the quality rating to scores on one of the environmental rating scales (such as the Early Childhood Environment Rating Scale, or ECERS), but some states do not require completion of an environmental rating scale to determine quality level, and some simply require that such a scale be completed to

inform program improvement (not tying quality level on the environmental rating scale to the level of quality within the QRS);
- While the broad domains of quality included in state QRS are similar, the number and content of quality indicators vary substantially across states.

Finally, the presentation noted that there are a number of important research questions that emerge from the widespread measurement of quality in state QRS. The research questions arise from different perspectives. For example:

- *From the perspective of implementation of a QRS*, key questions include the cost of ongoing data collection to assign quality levels, requirements for establishing and maintaining reliability across observers, whether self-report data included in the measurement of quality need to be verified, and whether designated levels of quality correspond to differences in observed quality.
- *From the perspective of parents,* key questions include how parents understand and utilize summary ratings and component indicators, and whether the availability of the information on quality affects choice.
- *From the perspective of providers,* key questions include whether the ratings are viewed as giving a fair portrayal of the quality of care they are providing, and whether the information is seen as useful in working towards quality improvements.
- *From the perspective of children,* the central question is whether higher levels of quality are associated with better developmental outcomes, across all children and for key subgroups.
- *From the perspective of the market,* the central question is whether the availability of information on quality increases demand for higher quality care, and whether this in turn has implications for the price of care and the supply of high quality care for families in different income ranges.
- *From the perspective of quality*, the central issue is whether there are statewide increases in the proportion of care that is of higher quality.
- *Finally, from the perspective of policy,* key issues are the cost effectiveness of QRS as an approach to enhancing quality, and whether there is constituent support for this approach.

In concluding, Zaslow and Tout noted the applicability of these issues to other types of quality initiatives.

Discussion followed the presentation. Among the comments and questions posed by participants were the following:

- State policies regarding the number and characteristics of providers who can remain unlicensed will play an important role in whether a QRS succeeds.
- While the presentation noted that the content and number of items pertaining to each aspect of quality differ across states, another key difference is the way in which different aspects of quality are weighted.

- In some states, policymakers and the business community are particularly interested in outcomes. They want to focus on the outcomes that QRS produce rather than on the inputs or components of quality. This requires a focus on what outcomes should be. For example, if a key outcome is school readiness, how does one define and measure this?
- Some states face the complexity of different measures of quality in different systems of early care and education, such as Head Start and pre-kindergarten.
- We need to use terminology that embraces all types of early care and education, including family child care. Quality rating and quality improvement efforts need to include all types of early care and education.
- It is important to acknowledge the constraints on parental choice: there is limited care available and limited transportation.
- The issue of diversity was not mentioned. Culture can be a very important component of quality and yet is not focused upon in depth in measures of quality.
- Addressing some facets of quality improvement may be more difficult than addressing others. For example, it may be easy to teach providers some things, such as how to plug electric sockets, but difficult to teach how to observe children, plan for implementation of a curriculum, or interact with children during ongoing activities.
- We need to focus on the facets of quality that are linked with positive changes in child outcomes. How much do we know about the things that need to be changed to really make a difference in child outcomes?
- Health and safety are foundations. They need to be ensured before focusing on other aspects of quality. They can be covered within a QRS or outside of it through licensing, but need to be addressed.
- It may be that no one measure of quality is sufficient. It would be extremely useful to have information to help with the selection of measures to couple with the environmental rating scales.

Plenary: What is the State of the Field in the Measurement of Quality

The next session turned to the issue of the state of the field in the measurement of quality. Lee Kreader moderated the session, and four presentations provided differing perspectives on the issue.

Donna Bryant focused on measures of quality appropriate for use in center-based settings. For this presentation, she reviewed the research about the functioning of measures of quality[1] that are appropriate for use in early childhood classrooms, including the Assessment Profile for Early Childhood Programs, Preschool Program Quality Assessment (PQA) (from High Scope), Caregiver Interaction Scale (CIS), Early Childhood Environment Rating Scale (ECERS), Infant and Toddler Environment Rating Scale (ITERS), Early Childhood Environment Rating Scale – Extension (ECERS-E),

[1] Child Trends has created an initial draft of a compendium of measures of quality in early care and education. The compendium will be released at the SAMI/CCRPC meeting (July 30 – August 2, 2007) and will be available online at http://www.childtrends.org or http://www.childcareresearch.org

Early Language and Literacy Classroom Observation (ELLCO), Observational Record (Ratings) of the Caregiving Environment (ORCE), Child-Caregiver Observation System (CCOS), Classroom Assessment Scoring System (CLASS), and Snapshot.

A key question in her review was what we want measures to accomplish.

- Bryant felt that a key criterion used to select and evaluate measures of quality in research is the ability to predict child outcomes with a focus on the use of factor analyses and factor scores that most succinctly define facets of quality that predict child outcomes such as early language and literacy development.
- One drawback to this approach is that the factor scores identified in this way tend to leave out the items pertaining to basic care, health and safety and the physical environment in facilities. However from a policy and training perspective, physical health and safety and basic adequacy of facilities may be high priorities. A key recommendation then is that factor analyses not drive all decisions regarding the key components of quality; priorities based on policy and practice should also be taken into account.
- Bryant noted that while a dominant approach has been the identification of measures of quality based on prediction to child outcomes, some measures have been developed with different criteria in mind. For example, the PQA was developed as a measure of the fidelity of implementation of a particular curriculum (High Scope). With increasing focus on intentional instruction in early care and education, this may become an increasing focus of measures of quality.
- In addition, the increasing prevalence of intentional instruction in early care and education has resulted in the extension of measures of quality from global or overall quality to measures focusing on the introduction of content, such as early stimulation for early language and literacy development. These are not so much measures of the fidelity of implementation of a particular curriculum, but measures of the extent to which specific content is being introduced and the quality of practices concerning instruction or stimulation in that content area. Examples of such measures are the ECERS-E and the ELLCO.

In addition to the content of measures, a key issue identified by Bryant was the way in which data on the quality of classroom environments are collected:

- She noted concern about the need to separate the use of quality measures to inform professional development, from research or accountability purposes. The same person should not, for example, complete an ECERS to help identify aspects of quality for the purpose of on-site coaching and also as part of monitoring for a quality rating system.
- Obtaining adequate reliability for use of quality measures in research or for accountability is a central issue. While it takes approximately one week to gain a basic understanding of the items in the ECERS and how they are scored, this is not sufficient to reach adequate reliability levels. Further practice is needed and should be accounted for in budgeting.
- The way in which reliability is computed is also important. For example, is it adequate to have agreement within one point, especially if the scale (like the CIS) is only a four point scale?

- The widespread use of measures such as the ECERS or the CLASS raises the central issue of what *systems* for training need to be in place. There should be explicit discussion of such issues as how much time and practice are needed, what role videotapes vs. live practice play, and who can serve as a trainer or an anchor for reliability assessments. While such systems of training exist for some measures, they do not exist for others, and this causes problems. Overall, the existence of a measure with good content and internal reliability does not assure that it can be implemented with good inter-rater and across-time reliability. Supports are needed to assure this.

Bryant then raised the issue of sampling. While sample size usually pertains to numbers observed or included in research, here the issue is twofold: both how many children need to be sampled, and what sample of time is adequate in order to get a picture of quality. She noted this as something on which studies differ and there does not seem to be consensus.

Bryant went on to discuss the research on what predicts our measures of quality. She noted an interesting difference by cohort of studies. Older studies have found structural variables, such as teacher education, class size and ratio to be good predictors of our quality measures. However more recent studies do not predict these measures as well as the earlier studies. Recent work, for example, finds teacher education not to be a strong predictor of quality. Bryant noted an emerging discussion in the literature on why this might be the case:

- Is it that the floor on measures such as teacher education has been rising, and so the range is narrower?
- Similarly, studies carried out in school systems, (more likely in the recent studies), may have a narrow range.
- Some research has found that teacher beliefs mediate the link between teacher education and quality. It is possible that there have been shifts over time in this underlying factor.

A further issue that is addressed in the research is the strength or magnitude of the association between quality measures and child outcomes:

- Bryant noted that while the major quality measures do predict child outcomes, we need to acknowledge that the prediction is modest; the associations are small.
- This may reflect limitations in the measurement of quality.
- However we also need to note that these associations have been found in correlational studies. A new body of research is focusing on efforts to improve quality, and examining child outcomes in light of particular treatments aimed at improved quality. We need to reexamine the strength of the link between quality and child outcomes when this new body of research is more fully reported.
- Sensitivity of measures of quality to such interventions is an issue also being addressed in the current research. Bryant noted that there is some recent evidence,

for example a recent report of changes on the ELLCO and ECERS in response to the Heads Up Reading program.

- Bryant pointed out that measures of quality were not included in some of the most widely referenced intervention studies, such as the Abecedarian and Perry Preschool projects.

Bryant then called attention to the issue of validity of quality measures across different cultures. There is some current research looking at the use of the ELLCO in classrooms with children learning English. There is also some current research looking at the ECERS in differing cultures.

Bryant concluded her presentation by reiterating that we should not overstate the strength of association between our measures of quality and child outcomes. The relationships are found consistently but are often small.

Questions and answers following Bryant's presentation focused on the number of children that suffice to provide a classroomwide picture.

- Bryant noted that 4-8 children are usually sampled. However some measures focus only on 1 child in the classroom (ORCE).

A presentation by Jean Layzer focused on measures of quality in home-based settings. Layzer began by noting that family child care as an entity exists on a wide continuum. Making further differentiations within home-based care is important, and Layzer suggested differentiating especially between care provided by relatives and non-relatives.

Unlike the measurement of quality in center-based care, for home-based care there is neither a roster of measures, nor a detailed literature to review that can result in the kind of overview that Bryant has provided. Key starting points for measurement of child care quality provided in home settings have been the FDCRS and the CC HOME. The FDCRS is closely related to the ECERS; it has very similar subscales. For a long time these measures dominated the field.

Layzer noted that Hispanic providers felt that these measures did not adequately reflect the quality of care they were providing; that the measures missed key components of what they were doing. There has been acknowledgment that measures of quality for home-based care do not fully reflect the broad spectrum of care that falls within this category.

Layzer noted that at the "turn of the century" (2000), researchers turned to the development of new measures, taking two different paths to expand the description of the quality of care provided in home environments. Porter and colleagues worked on the development of the Child Care Assessment Tool for Relatives (CCAT-R), a measure focusing specifically on relative care, seeking to capture the distinctive features of this type of care. The National Study of Child Care for Low Income Families (NSCCLIF)

developed the QUEST, a measure appropriate for use across all types of home-based settings (as well as center-based care).

The CCAT-R requires observation over half a day. It has five components: action/communication (20 items); behavior checklist (35 items); time-sampling; health and safety checklist (45 items); and materials checklist. The measure developed for the NSCCLIF involves checklists with indicators for health, safety, materials and resources. In addition, there are ratings of caregiver behavior, an activity snapshot, and child behavioral observation.

Layzer noted that the issue of prediction to child outcomes raised by Bryant was central and cannot be postponed. If we are investing money in improving quality, we need to see that children benefit and child outcomes are improved.

A key issue in home-based care is going to be intentional instruction. Providers in home-based settings sometimes say that intentional teaching is not what they are trying to accomplish. Observational measures confirm this. Is this an issue that is worth "going to the mat" about?

In questions and answers regarding this presentation:

- One participant noted that we need to be clear about what we are trying to build in home-based services. She noted that her state is updating its licensing for family child care. There was a surprise in the debate regarding introducing a requirement for education. At present there is no requirement for completion of a high school degree. The legislature questioned introducing an educational requirement. However the provider community supported the requirement.
- Another participant indicated that while child outcomes are important, outcomes for parents also need to be considered. For example quality of care has the potential to influence parental employment outcomes as well as parental mental health.
- A further comment noted that we should not assume that care by relatives is provided completely voluntarily.

Sharon Deich and Priscilla Little gave complementing presentations focusing on measures of quality in after-school settings. A number of issues complicate the measurement of quality in after-school settings.

- First, after-school care happens in a wide variety of places: libraries, museums, schools, community-based centers, sports fields etc.
- There is also a wide range in terms of the ages of children served (up to high school).
- In addition to these sources of variation, the "caregivers" themselves are more varied: ranging from volunteers to teachers, to coaches.
- Funding is also very different: funding is less stable (lots of time-limited grants) and involves fewer dollars, making it hard to plan in the long run.

- While parental choice is a major issue in early care and education, choice may be quite constrained in terms of after-school care. For example, in many instances if the children are in a particular school, that is where they are going to stay for after-school care.
- Children are only in after-school care for a few hours. We need to consider the implications of this for predicting from quality to child outcomes in these settings.

At present, there are a number of initiatives at the state level involving the creation of collaborations to focus on issues of quality in after-school care: how to improve policies and funding. Mott has such collaborations in multiple states. Parents are also very concerned about quality. Much of the discussion about quality in after-school settings focuses on academic outcomes, standards and licensing. There is some technical assistance going on with the aim of improving quality, and there are some studies currently evaluating efforts to improve quality.

Overall, about 6 ½ million children participate in after-school programs between kindergarten and 12th grade. There is a sense that more children would be participating if parents felt the programs were of high quality. Parents are paying attention to the issue of quality, with a particularly strong emphasis on this issue in middle school.

The first national set of quality standards for after-school care was released in 1998 (NAA). In 2002 Eccles led the work of the National Academy of Sciences on the key features of community-based programs. This work is seminal and is considered the "bible" for after-school programs. Eight (8) key features of such settings were identified and are widely used although with some differing versions. In a scan of after school program quality standards and assessment tools conducted for the Harvard Family Research Project, Priscilla Little uncovered 43 tools other than licensing and regulation tools to assess program quality[2].

A web-based poll was conducted to explore the extent to which programs are focusing on quality issues. This involved a small and clearly self-selected sample. The poll revealed that 1/3 of programs are not doing anything regarding program quality. Little noted that she felt that these findings understate the pattern: that at least half of all programs are not doing anything about quality.

A common criticism of quality measurement tools is that they are too broad, too complicated and too heavily focused on the provision of child care vs. the other types of activities after-school programs provide. They are not useful for reflecting on program quality in order to improve it. Very few of the existing measures are research-based or validated. Most tools are not age-specific, raising the question of whether different facets of quality are more important for children in the wide age range that participate in after-school programs. There is wide variation in how different constructs are included or

[2] *After School Program Quality Assessment Categories of Standards* available from the Harvard Family Research Project at www.hfrp.org

focused upon. Some constructs have emerged as more salient in recent measurement approaches. Examples include:

- Engaging families
- Program assessment
- Equity and cultural diversity

It is important to emphasize that there are no experimental research studies on efforts to improve program quality or on program variations, seeking to relate these to child outcomes. There is only one quasi-experimental study. The limited research suggests the importance of:

- The promotion of positive relationships
- Staffing
- Appropriate program structure
- Supervision

There are some studies in the field that should begin to tell us more about child outcomes.

Questions and answers following this presentation focused on the issue of child age. It may be possible to keep the same construct but modify the indicator according to child age. Discussion also focused on parental involvement and the need for flexibility in how parents can become engaged. There is a need for active outreach to help parents understand the importance of their involvement.

Tamara Halle then provided an overview of a Compendium of Measures of Quality being developed by researchers at Child Trends. She noted that the purpose of the Compendium is to provide a consistent framework within which to summarize information on existing measures of the quality of early care and education settings. This information will serve as a resource to researchers as well as practitioners, to inform the development of professional development systems and for the measurement of quality for policy-related purposes.

Two products are currently being developed: (1) a compendium that will provide information on the characteristics of measures of quality, and (2) a matrix that will summarize the constructs covered by each quality measure.

The compendium is modeled after the Early Childhood Measures Profiles[3], created for the SEED (Science and the Ecology of Early Development) consortium of federal agencies with a focus on early childhood development, which summarizes

[3] The Early Childhood Measures Profiles can be accessed through the following websites:
http://aspe.hhs.gov/hsp/ECMeasures04/index.htm
http://www.childtrends.org/files/earlychildhoodmeasuresprofiles.pdf
http://www.childcareresearch.org/location/ccrca8634

information on the reliability and validity of widely used assessments of early childhood development.

Profiles of 33 measures of quality are currently being developed for the Compendium of Measures of Quality. Each profile in the compendium includes:

- Background information: author and publisher; the purpose for which the measure was developed; population, age range and settings the measure is appropriate for; focus on diversity as a component of quality; and key constructs covered. The profile also provides reviewer comments about the measure's background.
- Administration: training required for administration of the measure; time needed to administer the measure; and cost of the measure.
- Functioning of the measure: information on the reliability and validity of the measure.

The purpose of the matrix is to compare the constructs covered in each quality measure to domains of quality deemed important to the creation and functioning of quality initiatives. It is also intended to enable the examination of possible gaps between existing measures and what a quality initiative may be evaluating. The domains of quality included in the matrix are:

- Relationships
- Curriculum
- Teaching
- Assessment
- Health and safety
- Physical environment
- Teachers
- Families
- Community
- Leadership and management

Halle concluded the presentation by emphasizing that the feedback of participants at the Roundtable was important. She requested input in terms of measures that should be either added to or dropped from the compendium. She also requested input on the domains of quality used in the matrix. She asked if the summaries that each of these will provide are likely to be useful to multiple audiences.

In questions and answers regarding this presentation:

- One participant asked whether it would be possible to include in the matrix additional dimensions focusing on policy, cost, and professional development.
- The question was raised of how English language learners and children from culturally diverse populations will be included. Dr. Halle noted their inclusion in the matrix.

- The importance was noted of including measures of directors, not just teachers or caregivers, in measures of quality.
- It was suggested that introductory material be included that notes that individual measures were sometimes developed with the intent that they be used in combination with other measures. As a result, they may have strengths in terms of coverage of some constructs and gaps for others. Without such a statement, there could be unintended negative consequences of the matrix, such as apparent negative evaluations of the coverage of measures.
- New measures are being developed focusing on stimulation for early math, and also regarding physical activity/obesity. It was suggested that these measures be included in the compendium.
- Extent of coverage of particular constructs (in terms of number of items used to cover the construct) was noted as potentially important to record.
- Similar efforts have been carried out within specific states. The suggestion was made that the Child Trends team review these to be sure the compendium is not duplicating previous efforts.

We discussed plans for the following day and adjourned for the day.

TUESDAY, DECEMBER 5, 2006

Tuesday morning and afternoon each included an orienting plenary session, followed by concurrent breakout sessions and reporting back of key issues from the breakouts to the full group. The morning focused on the measurement of quality in state initiatives, while the afternoon focused on research perspectives on the measurement of quality. In addition, over lunch, a panel presented on federal research initiatives with a focus on the measurement of quality.

Morning Plenary: Measuring Quality in State Initiatives

Judy Collins noted guidelines for the Plenary on Measuring Quality in State Initiatives. Again, quality ratings systems (QRS) were selected to illustrate state quality initiatives. Representatives from four states were asked to give presentations of about 10 minutes each focusing on the history of their state's QRS and noting the current status of their initiatives. Each state representative was also asked to pose one question for the researchers in the room. After the plenary, the floor was opened for questions.

Jennifer Towell presented on the development of a QRS in Oklahoma. The QRS in Oklahoma is called Reaching for the Stars. It was established in response to a state task force focusing on how welfare reform would affect child care. The objectives are: to increase the overall quality of child care; to provide guidance to consumers in evaluating programs; and to increase the availability of high quality care and overall slots for child care through the use of higher reimbursement rates for higher quality care. The QRS was launched in February 1998 with two levels: 1 star indicating that minimum licensing requirements have been met, and 2 stars indicating that additional criteria were met. Two further levels were introduced later: a 3 star level and a 1 star plus level (added because the distance between minimum requirements and the two star level was too large for some caregivers).

- The 1 star plus level has a requirement for additional training, a learning environment component, daily schedule component, and a parent involvement component (e.g., parent-teacher conferences, parent input on policies).
- The 2 star level involves either national accreditation OR all of the qualifications of 1 star plus with the addition of specific educational requirements, use of an environmental rating scale as a tool for self-evaluation, development of improvement goals, staff compensation that reflects experience and professional development, and parent and staff surveys.
- The 3 star level involves national accreditation AND the further criteria noted as alternatives to accreditation in the 2 star level.

Of approximately 5,500 licensed facilities in the state, about 3,000 are in the star system. Participation is voluntary except when programs want to contract with DHS to care for children receiving subsidies (in which case they must have a 1 star plus level or higher). Facilities can lose their star ratings, though the rate is very low (about 2 percent). The two most common reasons are not meeting minimum licensing requirements (e.g.,

for serious lack of supervision, etc.) or a facility remaining in level 1 plus for a period of longer than 24 months (this is seen as a temporary rating, and the facility must be ready to move up to a level 2 at the end of 24 months or will need to fall back to a 1).

There are five assessors and five outreach specialists (one assessor per specialist). They carry out assessments only in 2 star facilities, as 3 star facilities have accreditation and are not considered to need assessment. The environmental rating scales are not used in the assessment process. They are used instead as a tool for programs to set goals for improvement.

The question posed to researchers at the end of this presentation was: In Oklahoma, facilities are monitored through the licensing system and are visited three times a year. What criteria could be observed through this process that are definable and objective?

Paula Neth presented on the development of the QRS in Colorado. Educare Colorado was created in response to the findings of the Cost Quality and Outcomes Study results indicating that 87% of center-based care in Colorado was of "poor" or "mediocre" quality. In response, the Governor created a task force, including business leaders, politicians, and practitioners. The task force identified the need for an accountability system that would inform parents about quality in child care and provide a framework for improving quality. Educare Colorado was created with the needs of three groups in mind: parents (how to help parents identify quality in an easy way); government (providing accountability for funding on quality); and practitioners (a system that honored all types of early care and education so that all could progress in terms of quality).

In 2002, ratings were carried out in 6-12 sites as a pilot research project. School readiness legislation passed that year helped to broaden the outreach of the program. Further helping to extend the program's outreach into a statewide system, Educare Colorado merged with the state's Child Care Resource and Referral (CCR&R) Agencies. The 15 CCR & Rs deliver the quality improvement work aligned with the ratings. The QRS in Colorado is a voluntary system implemented outside of the government.

The rating system has five components, each of which (with the exception of accreditation) can provide up to 10 points toward a total score. The components are (1) a rating of every classroom using the ECERS-R, with scores averaged across classrooms; (2) training and education; (3) ratio and group size; (4) family partnerships; and (5) accreditation. The total number of points translates into the number of stars, with the number of stars ranging from 1 to 4. At the completion of each assessment, the provider receives a quality performance profile that details the strengths in each program and gives recommendations for quality improvement.

Questions now being asked about the implementation of the QRS in Colorado are: (1) whether environmental rating assessments are needed in every classroom, which is very costly, or if a sample would suffice; and (2) how to connect the QRS to the licensing system in the state. Early on, it was decided that the QRS would be administered outside

of the licensing system in Colorado. This in part reflected concerns about the heavy caseload of licensors (300) and the fact that most programs are not visited annually. There are currently efforts to pass legislation to require annual visits. There are concerns about programs with health and safety issues.

Two related questions posed to researchers at the end of the presentation were: What quality improvement and technical assistance approaches are most effective? Given limited resources, where do we get "the biggest bang for our buck" in terms of interventions to get children ready for the classroom?

Betsy Farley presented on the development of a QRS in Kentucky. The QRS in Kentucky is called the *Stars for Kids Now* program. Eleven years ago, the Governor became very interested in the research on early brain development. An early childhood office was established within the Governor's office, and developed a 20 year comprehensive plan. The Stars program grew out of that plan. It was implemented statewide in 2001. This is a voluntary program in which about 1/3 of the state's licensed or certified centers and family child care homes participate. The state licenses child care centers. It also licenses homes that care for up to 12 children. It certifies homes that care for up to 6 children. All three types of settings can participate in the Stars program.

There were a number of goals for the program from its inception: It was seen as needing to build on a broad consensus of participants, and to involve coordination across multiple programs/agencies, including the Department of Community-based Services, the Department of Public Health, and the Department of Education. In addition to coordination across programs, coordination with standards for children when they enter school was viewed as a priority. Continuous quality improvement was an important goal, involving the provision of technical assistance by qualified staff. It was a priority to develop a QRS that could be communicated easily to the public, and to take the steps to educate the public about what quality is.

The program aims, through its different components, to:

- Increase parent involvement;
- Reduce ratios;
- Improve the quality of children's environments above licensing level (Stars means above licensing);
- Improve professional development, encouraging staff to complete college programs as well as to complete the Commonwealth Child Care Credential (half way to the CDA; seen as a stepping stone on the way to a CDA);
- Encourage retention by requiring increased compensation;
- Provide incentives according to the number of subsidized children.

A number of adjustments have been made to the program. These include:

- The establishment of growth targets;
- Centralized administration and management of the program;
- Contracting out for the provision of technical assistance;

- Refocusing of the technical assistance efforts beyond basic licensing standards to higher levels of quality, and to keep working with centers or homes so that they aim for the higher quality levels over time;
- As of October 1, 2006, if centers care for 40 or more children receiving subsidies, they must work towards and stay in the star program in order to continue to receive subsidies. The goal is to provide supports so that they succeed within the Stars program, but receipt of subsidies will be discontinued if the center or home is not progressing in a six month period;
- Looking at how often to schedule visits.

The question that Kentucky would like help with from researchers is how to address the demands of policymakers for data on child outcomes; that is, how to determine whether children are benefiting from the investments in quality made by the state in terms of their school readiness.

Jamie Gottesman gave a presentation on the development of the QRS in Ohio: Step Up to Quality. Planning for the QRS in Ohio took place in 1999 and was funded by Sisters of Charity and other foundations. Judy Collins worked with a group of key stakeholders. However, work towards implementing the system had to be shelved during a period of cost containment. The fact that a plan had been developed proved useful later when it was found that Ohio had not spent the minimum required for the 4 percent set aside of the Child Care and Development Fund. When the resources became available, because the plan already existed, they could move forward with it. In 2004, the plan was reviewed in light of new information and steps were taken to begin its implementation.

The design of the Step Up to Quality involves building blocks above the licensing level. Step one is something that providers view as achievable. It is an "awareness" step focusing on things providers should be thinking about. Step two involves the application of some of the new awareness. Step three involves change throughout the care setting.

The components of the QRS focus on:

- Ratio
- Staff education
- Specialized training
- Administrative practices
- Early learning

In identifying specific indicators to operationalize these, it was viewed as important to identify a list of essential elements rather than an exhaustive list, in the hopes that requirements would not be overly burdensome and to encourage participation. It was also viewed as essential that indicators could be observed (rather than inferred), and that each indicator was grounded in research. A high priority has been placed on indicators that reflect differences in levels of quality and yet are not burdensome to collect data for.

A pilot study was completed in 2005. It found the average environmental rating scale at step one to be 4.6. The average at steps two and three was the same at about 4.92. An alignment document has been developed to relate step three in Step Up to Quality to NAEYC accreditation and to the Head Start performance standards. The pilot study shows steps two and three to be very close to the "good" level on the environmental rating scales. This is seen as a strong foundation on which to build, for example for pre-kindergarten programs.

The question posed for researchers in this presentation was: It may be more important to consider a combination of indicators as determining levels of quality rather than individual indicators. How can we identify the mix or combination of indicators that would be most effective in distinguishing levels of quality?

Questions posed by Judy Collins to presenters. At this point in the panel, the moderator posed a set of questions to participants.

A first question asked whether the QRS in each of the states had requirements involving participation by providers serving subsidy-receiving children.

- In Oklahoma, participation is mandatory for programs that want to contract with DHS to care for subsidy-receiving children. A large majority of centers and family child care homes with such contracts are participating in the QRS (over 90% of centers and over 70% of homes)
- Kentucky found that this was not happening spontaneously and so made participation mandatory for centers caring for 40 or more subsidy-receiving children.

A second question asked about what incentives states use, especially to encourage participation by those caring for low-income children.

- Each state described incentives for providers serving children from low-income families. For example, in Ohio, funds are available based on star rating and number of subsidy-receiving children for making repairs to facilities, for curricular tools, or working towards accreditation.

A third question asked whether states are looking at the geographical distribution of programs at higher program quality levels to see if they are in areas where low income families live.

- Oklahoma's program is statewide and so serves all 77 counties. There are 3 metropolitan areas with high concentrations of low income families that have high participation rates.
- Kentucky tends to have better success in getting programs in rural areas to participate in the QRS than in other areas.

Morning Breakout Sessions

Four breakout sessions provided the opportunity to have in-depth discussions regarding the measurement of quality within state initiatives. Each moderator was asked to bring back to the full group "takeaway messages" from the breakout session. We summarize key themes as well as the takeaway messages from each session.

Breakout on Aligning the Measurement of Quality with Other State Initiatives. This breakout session was facilitated by Beth Rous. Susan Neuman served as discussant.

QRS are moving forward simultaneously with the development of other standards, such as early learning guidelines, core competencies/bodies of knowledge for practitioners, Head Start and pre-k standards and accreditation standards, as well as systems for professional development. Licensing and regulation also play a critical role in establishing lower thresholds for quality. This breakout addressed the question of how work across these systems can be coordinated.

How does your State QRS or other quality initiative incorporate such components as early learning guidelines, core competencies, accreditation and licensing, pre-K, and Head Start?

The participants of the breakout group spent much of the time discussing what elements of the system should be aligned, and how. There was agreement that alignment needs to be systemic, strategic, and measurable. What emerged from this exercise was a conceptual model of an ecological system that had child and family outcomes at the center, and concentric circles that include the following:

- Adults who interact with children
  - Caregiver-child interactions
  - Caregiver knowledge and behaviors
  - Parent education
- Programs
  - Administrators
  - Staff qualifications
  - Turnover
- Systems
  - Standards
  - Financing
  - Supports for professionals
  - Communication outreach

It was noted that Zero to Three has a similar model representing key elements of the ECE systems that support quality care for babies and toddlers, but that model has five concentric circles around the child rather than the three articulated by this breakout group. The group noted that they liked having just three concentric circles because they felt you should never be more than two steps away from the child and family outcomes which are central to the system.

The first question that should be asked is "what are the outcomes we want for programs, for families, and for children?" The second question should be "how do we build the systems that are needed to achieve those outcomes?" All levels of the model need, ultimately, to link to child outcomes. At the core, you have to have a vision of what you want for children and families, and try to pull the pieces together around that vision. In every state, the elements will be different. Each state could really answer some of these questions about synergy.

One participant noted that she thought the model needed to be three-dimensional, and a discussion of the usefulness of logic models to show directionality of linkages took place. A logic model is a roadmap. It can indicate the necessary ingredients on one side, and show what is most powerful in reaching child and family outcomes. It can also show that there are multiple mechanisms to reach a common goal. It was decided that both a conceptual model and a logic model were important, and that they needed to work together. A logic model is useful in that *the exercise of developing the logic model* helps all those who built and use it. However, some argued that a logic model is not as helpful to policymakers because they do not think that way.

An important point was made about the term alignment. Alignment does not mean "the same," but rather, it means that there is congruency across systems. When we are talking about congruency we are saying that one part of the system does not conflict with another.

The systems within a state that need alignment include:

- Professional development systems
- Quality rating/enhancement systems
- Licensing
- Accreditation
- Career lattice
- Head Start/Early Head Start
- After school
- Infant/toddler
- Special education
- Early learning guidelines
- State pre-K

<u>What successes have you had in your State with the coordination of systems?</u>

It was suggested that having a single body within the state that has oversight and coordination of multiple ECE systems (e.g., professional development, licensing) can be helpful. However, even if a body exists within a state for coordination, it might not have the authority to make decisions about alignment. Beth described activities in Kentucky where the state administrator and staff from the Department of Education were working together so that QRS for centers mirrored the system for State Pre-Kindergarten. One lessoned learned from this experience is that it is important to have all agencies at the

same table. In Kentucky, bringing everyone to the table was a key step in authority being provided for oversight of a 20-year plan. They are using tobacco settlement money for funding, leveraging resources to address gaps. It is too early to see improvements in child outcomes because the assessment system is just being put in place. They think that children are in a better place, but they do not have data on child outcomes yet. The outcomes for subsidized children will be a key focus.

States are developing professional development systems with career lattices. How important is it that these lead to qualifications for different levels within QRS? States have moved forward with different initiatives with what was known at that time. Now there is a need for agencies in charge of different ECE systems to talk to each other. An example was given of a state that had developed its QRS prior to substantial work on a career lattice for professional development. Now there is a challenge of needing to go back and link the QRS to the new information in the career lattice.

What have been the most significant challenges with the coordination of systems in your State?

One of the most significant challenges to coordinating systems is the limited resources available. Given limited resources, how do you make the pieces work together better? The importance of blended funding for the purpose of alignment was discussed.

Another challenge states have faced is reaching consensus on standards. Reaching consensus on a state's Early Learning Guidelines is often not difficult, but reaching consensus on program standards is much harder. For example, the Head Start standards do not make sense for home-based programs and therefore cannot be adopted wholesale for another setting. There is discontinuity in expectations and standards that apply based on funding and setting. (A big issue is talking to providers with multiple sources of funding or children in multiple places over the course of a day.) A participant suggested that perhaps the insight is not to require shared program standards but rather to have each program think about keeping child and family outcomes at the center of their goals when creating their program standards. States can say we don't want the same thing; we just want your standards not to conflict with one another.

Lack of political will to create alignment was identified as a challenge. Another challenge is that many different things are going on at many different levels. For example, at the same time that we have technical assistance focused on programs, we have educational requirements for individual educators (e.g., 50% with AA or BA in Head Start), and a mentor coaching program. We need to acknowledge that many things are happening at the same time. How do we assure that those who are getting an education are actually changing their practice? How do we determine the cumulative effect of all of these activities at the program level and level of the individual early childhood professional on child outcomes?

Another challenge is not having good measures to determine alignment across systems. But to have good measures, you must first be able to define what you want to measure. If you don't measure it, it doesn't get valued.

What tools or processes would be most useful for states to incorporate to create more complete alignment?

Participants felt that the best process for establishing alignment is to make sure all stakeholders are at the table from the beginning. Multiple agencies need to be involved. There are states that are bringing everyone to the table for new initiatives such as a kindergarten readiness assessment; a QRS. That is what a discussion of alignment is most helpful with.

However, in addition to collaboration, participants felt that it was also critical to have leadership. There must be a lead agency or person to spearhead the alignment activities and make sure they move forward.

Participants felt that it was also important to have information about the implementation of systems and standards in the states. A lot of implementation is going on (e.g., implementation of Early Learning Guidelines), but it is not in the public eye. Most states are not recording what they are doing, and it is not scaled up to a statewide level. How do we get this information out? No one is writing about it. Research is so focused on the classroom and interactions; we also need to be looking at implementation issues at this broader level.

On a related note, participants in the breakout session felt that information about standards needs to be better communicated through outreach to program directors, providers, and the public. There are 49 states that have Early Learning Standards. We have no idea if teachers have heard about them.

The latest research findings should inform the work in the states. For example, Bob Pianta's research suggests that teacher interaction with the children is what really matters. He unpacks what we mean by intentional teaching[4]. Now we need that research to inform what we do in teacher preparation. A participant spoke of a two-year study which found that the typical coursework caregivers took as part of professional development had virtually no impact on practice or child outcomes. What was critical was coaching. A very intensive coaching model can be effective in changing practice and child outcomes. How do we begin to get some of these key factors into the system? One vehicle is to bring researchers and policymakers together around these issues of quality, professional development, and alignment of systems.

Another participant warned that we should not focus exclusively on what goes on at the classroom level. When we talk about research on professional development, we are talking about research at the classroom level. But a lot of what is really important is at the

[4] See for example Pianta, R.C. (Winter 2007). Preschool is school sometimes: Making early childhood education matter. *Education Next* (1). http://www.educationnext.org

director level. We should consider the importance of consultative technical assistance to directors. There is tension between our levels. How is TA allocated and who gets it?

Others agreed that the role of the administrator is critical. The administrator makes sure there are good relations with families; that there is reflective practice happening in the classroom; and that there is outreach to the larger system.

Alignment of data systems is important. If we can't link data about professional development to the QRS, for example, it is problematic. As states are thinking about measurement, let's bring measurement people together across the agencies. This will make a difference.

It was noted that small experiments within states can be an important enhancement. For example, a study of fidelity of implementation can be carried out within a small scale experimental study to confirm that there are changes in classroom practice in response to an intervention to improve quality. Why go all the way out to child outcomes if you don't first confirm there have been changes in practice? Ask first if adult behavior changed. The political pressure to have answers about child outcomes can be met if you take the approach of small experiments.

Another idea was to focus on getting alignment between one or two elements at a time, rather than trying to align all systems at once, which can be overwhelming. The message to states should be that there are small things you *can* do. Gwen Morgan's idea of a three-legged stool is relevant here. You can't do anything to one stool leg without affecting another.

Summary of Take-home Messages:

- Alignment does not mean that everything is the same. Rather, it means congruence across systems.
- Alignment efforts need to be strategic, systemic, and measurable.
- Alignment requires an ecological model with child and family outcomes at the center.
- A logic model is helpful to articulate directionality and multiple paths to the same goal(s).
- The ecological model and the logic model need to work together.
- Stakeholders need to agree on definitions of constructs that are being measured for alignment across systems.
- More attention needs to be paid to the importance of program administrators because they are connected to all levels of the alignment model (parents, teachers/classrooms, programs, systems).
- Bring researchers and policymakers together.
- Collaboration across multiple agencies is needed for alignment within states, but leadership is also needed.
- Linkages of program data systems would facilitate alignment and accountability efforts.

- Documenting implementation is important and necessary before you start to look at outcomes.

Breakout on the Components of Quality in State Quality Improvement Initiatives. This breakout session was facilitated by Marilou Hyson, with Mark Lewis and Jamie Gottesman serving as discussants.

The breakout session focused on the key differences across states in the measurement of quality in state quality initiatives. While many components of QRS are common across states, significant variations exist in the breadth and depth of domains covered. The thresholds for quality also vary widely, with a broad range for the highest level of quality. What are the key distinctions that can be made across states? What are the implications of state variation in the definition and measurement of quality? What are the challenges of communicating these variations to the public, and in particular, to parents?

The discussion started out with Mark Lewis providing information on the QRS in Oklahoma. Overall, he feels that the QRS has contributed to helping providers increase the quality of care that they give. One of the challenges that Oklahoma faced in terms of building their QRS was engaging parents. It has been much easier to include providers but the parents' input is also critical. Another challenge has been focusing on interactions and not just on the physical characteristics of settings.

Jamie Gottesman shared her experiences from Ohio, agreeing with Mark Lewis that people's perceptions of high quality often focus on physical/ environmental aspects of a program and overlook interactions. In Ohio, all licensed providers participate in the QRS. The system consists of five benchmarks and three levels, ranging from an awareness of selected indicators, to a level of planning how to improve on these various indicators, and finally to implementing changes. Early Learning has just been added as a benchmark in Ohio. Gottesman raised the question of whether improvements in teacher qualifications and professional development can be measured solely using environmental rating scales.

The group went on to discuss differences across states in including parental involvement in their quality rating systems. While some states choose to include parental involvement within other components, other states choose to let parental involvement be a stand-alone component of their systems. It would be interesting to find out what criteria states use when making a decision about parent involvement as a separate component or embedded within other components.

What states felt was helpful in putting together their quality rating systems was having access to quantitative and qualitative data before implementing their systems, having several statewide people and departments all working together, learning from and working with states that came before them, and having previous experience with technical assistance systems and environmental rating scales.

To the question of how health and safety are included in state QRS, states answered with a range of options that they had chosen. Whereas some states have included it in their basic licensing requirements but not on a higher level, others have included it as a whole separate component, and yet others rely on environmental rating scores to address this issue. It was noted that the issue of health and safety as an indicator of high quality versus a basic requirement that should be a given across all programs was one of the controversies in the revision of the NAEYC standards.

The take-away messages from this breakout session were:

- The process of agreeing on components typically involves more than looking at the data. Identifying a small set/list of components of quality is more complicated than it seems and involves unpacking complex issues, such as issues of privacy and turf issues.
- How health and safety fit into state systems is still a complex issue. Some states view health and safety as a component of its own while others feel it should be embedded within other components.
- Lamppost issue (the light will shine only where there is a lamp) – There are certain components that are harder to measure than others – inherently or conceptually. If we have a tool to measure a certain component of quality it will get measured, but if not, the component will take a back seat.

Breakout on Balancing Documentation, Observation and Verification. Anne Mitchell facilitated this session. Deborah Neil served as discussant.

States vary in the emphasis they place on self-report, collection of documentation, review of materials, and inclusion of site/observational visits in their quality improvement initiatives. There are also variations in the content and depth of information that is required to collect or observe. This breakout session focused on the issues of where states are putting their resources for the collection of information about quality, and what they are learning about the value of different sources of evidence.

Deborah Neil started the session by talking about the different types of data collected as part of Tennessee's QRS. By way of background, in Tennessee, any facility serving more than 4 children must be regulated, and 80% of subsidy children are in regulated care. There are no set ratios in Tennessee, but a set group size. The QRS originally got funding due to the tragic deaths of children in child care vans. That started a big movement that led to legislation.

The QRS system has three levels, and these are decided upon based on seven components. Six of these components are structural components (including ratios, staff compensation, professional development, on-site technical assistance, and college courses). A licensor evaluates these components. Each licensor has a caseload of about 35 centers/providers. They do 6 unannounced inspections per year, including in family child care homes.

Additionally, assessors conduct the environmental rating scale. Deborah Neil noted that the QRS in Tennessee places great emphasis on on-site evaluation. Every program gets evaluated annually using environmental rating scales (ERS). The ERS is the only learning environment tool currently used, but they are exploring other options with researchers at the University of Tennessee.

Providers form professional development plans for continued education. In Tennessee, college courses must be "seated" not online. CCR&R staff does a lot of on-site assistance to help with professional development plans. The Family Child Care Alliance also helps family child care providers with thinking through and achieving professional development plans and goals. Family child care providers receive help with business plans to help them manage the money and business aspects of running their business.

Participants from other states illustrated the range of documentation approaches. In one state in which the QRS is just rolling out, much of the documentation is provided through self-report. There are voluntary classroom trainings on FDCRS, ECERS, etc. Health consultants assess providers at certain levels and send in documentation. Any training has to be approved, and CCR&R's conduct and document trainings for providers. However for educational attainment, transcripts are not required. The emphasis on self-report in the system is intentional, but some are raising questions about the reliance on self-report without verification for some of the information.

One participant noted that verification is not just an issue for information provided through self-report but can even be an issue with respect to on-site observations. She noted an experience in a state in which providers would bring in additional teachers and materials for their observations (which were announced). Providers were probably going to great lengths to get better scores on their assessments because these are high stakes assessments which could lead to better incentive payments. It is understandable why NAEYC does unannounced visits.

A participant from another state noted that they have really struggled with what data to rely upon. Some issues have come up with self-report regarding the documentation of early learning indicators. There was some tension with family child care providers when asked to write out their curriculum: they did not want to be scored the same as centers, and evaluators struggled to score their curriculum/action plans.

Another state just launching its QRS is trying to do a combination of documentation and observation. They feel that self-reporting is beneficial for children because it requires the provider to reflect on the care he/she is providing.
An interesting model that this state is focusing on comes from nursing homes in which there is a combination of observation with self-report documentation that can be audited.

Participants in this breakout decided on the following "takeaway messages:"

- There is variation in assessment approaches in different QRS systems, ranging from systems that focus heavily on observation, to systems that are based on a combination of the provision of documentation and observation, and to those heavy on self-reporting and the provision of documentation.
- QRS assessments are a challenge because assessments are high-stakes due to incentives.
- It would be valuable to explore the nursing home model of observation with documentation that can be audited.
- There is recognition of the need to balance the money invested in a QRS across the cost of the assessment system, of providing further professional development, and providing incentives for improving quality;
- Self-reporting can be a useful tool for providers to improve quality through self-reflection.

Breakout on State Research Needs. This breakout session was facilitated by Craig Ramey. Michelle West served as discussant.

State quality initiatives and systems cannot wait for research to inform every decision that is made. Yet is useful to identify the research questions that state policymakers and program administrators feel are critical for informing their systems and programs. The breakout focused on the issue of which research questions are most pressing. It also asked where the research base is strong enough to draw from now and where we urgently need new research efforts.

Dr. Ramey opened the session using the data collection for Louisiana's state-funded pre-k (LA4) to illustrate key issues related to state data collection. This is a program initiated six years ago using TANF and other federal funds. Implementation started with a pilot program in 11 parishes (counties). The pre-k classrooms are in public schools, have certified teachers, have a maximum class size of 20, ratio of 10:1, and utilize a uniform curriculum. It also includes health care, wraparound services and summer programs. Louisiana is looking at using these programs as sites to provide family health care. The program is free for children from families with incomes below federal poverty but there is tuition for those from families above poverty.

In terms of data collection, ECERS observations are carried out annually and scores average 6 over a period of years. Teachers complete the Developmental Skills Checklist when the children enter and exit the program. In the first year, children exited the program with scores that averaged at the 35^{th} percentile (N=1300). In the second year, scores increased substantially, and children existed at between the $50\text{-}59^{th}$ percentile. The greatest gains were made by lower income children.

Dr. Ramey noted that the universal implementation of the program makes it impossible to collect data from a control group. Instead, in addition to ECERS and exit scores on the Developmental Skills Checklist, they have looked at data on grade retention and use of services for special needs for children who have and have not participated in LA4, looking also at subgroups who are and are not participating in the school lunch

program. Children who participated in LA4 had a 7 percent chance of grade retention while those without LA4 had a 12 percent chance. In addition, Louisiana children do better on test scores when they have participated in these publicly supported pre-k programs than those who do not, and the difference is especially marked for low income children. Further work is taking place now looking at growth trajectories over time of children who have and have not participated in the program.

Dr. Ramey noted that states want data that can help indicate whether a program is worth investing in/is producing the results it is looking for. He noted the importance of states investing enough funds in a system of monitoring and assessment to be able to make this judgment. He suggested a system of ongoing universal assessment of children as an efficient assessment approach. When asked what he felt was the key to the positive results from this program, Dr. Ramey indicated that the teachers were the key.

In the discussion that followed, a number of themes emerged:

- It was agreed that states are interested in both short and longer-term outcomes of participation in early childhood programs. Participants discussed whether participation in a program in the pre-k year should be seen as responsible for third grade reading scores.
- Participants noted that states are beginning to question whether measures of program quality like the ECERS are enough to inform states about investment of funds. There was agreement that states want more. Participants pointed to the importance of the approach of collecting data on the quality of environments and on child outcomes, and relating these.
- Yet at the same time, not all states are looking at child outcomes. One participant noted that her state is focusing first on the foundation of safety in programs before turning to the issue of school readiness; building up from the foundation first.
- Participants noted that when child outcomes are examined, there is often not good alignment between what is actually taught in the early childhood program and the content of the assessment. Alignment should be a greater focus.
- Participants noted that children are not the only ones who may show outcomes from participation in early care and education programs. Adults may benefit in terms of employment outcomes. We need to look at both adult and child well-being outcomes.
- Participants asked what the key features of quality are that are necessary to produce positive child outcomes for 4 and 5 year olds. They noted the following as key indicators: interaction, health and safety, curriculum, and administrative leadership.
- Participants noted that governors often have a four year “window” in which they want to accomplish something. But it may take longer to accomplish something important in a thoughtful way.
- There was discussion about competition within states for funding to go to research vs. to run the program (e.g., staffing, paying higher wages so that there is retention). There is also a central issue of capacity within states to carry out

research. There is a need for expertise within states to do evaluations correctly; a need for state policymakers to be able to call upon research expertise more easily.

- The discussion also focused on a tension between ongoing tracking and monitoring of program quality and child outcomes over time, and doing more focused or modular evaluation studies.
- A key issue is also program dosage that children receive. Low income families move around. Children may not participate in a program for a full year.
- There are central issues regarding understanding the multiple programs or policies that families participate in. For example, we cannot identify which children participate both in Head Start and TANF. States are working on using a unique identifier so that it is possible to link data from different administrative systems.
- There was a discussion of the relevance in one state of experiences or data from another. States want to have a picture about what is happening in that specific state. It may not be necessary to get complex data to get a picture of what is happening within the state, but state-specific data are a priority. There was a discussion about focusing on a limited number of core issues rather than trying to present too much to legislators or the public.
- One participant noted that it is not enough to have the data; it is also important to think about analyzing and reporting to the public on a regular basis.
- Participants suggested connecting researchers in different states, seeing if it would be possible to address similar questions and build on each others' experiences. Researchers from state universities could form a consortium. There are models within states of such collaborations.
- An issue of particular concern was whether low income children are participating in the programs that are high quality, or whether they have difficulty accessing high quality programs. One participant noted the importance of monitoring strategies intended to increase access of low income children to high quality programs.

The three takeaway messages that the group decided upon from this breakout session were:

- The idea of developing a consortium of researchers (within states as well as at the national level) to work together on this set of issues, where possible addressing common research questions and learning from each others' experiences.
- The need for research linking process variables to child outcomes; not just focusing on outcomes.
- The need for data infrastructure to track children over time. There is a need to follow children across time and not just pick them up when they are 4 years old.

Lunch Conversation: Points of Intersection With Federal Research Initiatives

In this conversation, staff from federal agencies was asked to reflect upon issues related to federal research initiatives. Colleen Rathgeb (Office of the Assistant Secretary

for Planning and Evaluation) moderated the session and five presenters shared their perspectives and their experiences with the audience.

Mary Bruce Webb shared the perspectives of the Office of Planning, Research, and Evaluation, DHHS. Many of the quality issues that have been raised during the meeting have been talked about in the context of preschool and child care programs, but Head Start Research is facing many of the same issues. They have been looking at multiple levels of service and multiple dimensions of quality in Head Start (characteristics of families, quality at the classroom and provider level, etc.), but it has been difficult to get at the less tangible issues, such as teacher-child interactions. They have not been able to successfully measure that with the existing measures. Another point in which they are very interested is the structural and organizational supports that sustain quality in Head Start, as well as the community level factors that support quality (e.g., resources, policies, expectations at the community level).

Mary Bruce Webb further expressed a concern of Head Start research not attending to culture as much as needed, in terms of quality. This is a particularly important issue in Head Start, where 65% of children come from minority backgrounds, 13% are children with special needs, the number of English Language Learner (ELL) children is steadily increasing, and there are a number of other special populations, such as migrant farm workers, that Head Start serves.

James Griffin provided an overview of research initiatives within the National Institute of Child Health and Human Development. NICHD funds numerous research efforts, from basic research to field research. One of the historical studies to show that early child care does not have a negative impact on children was also funded by NICHD. Currently, NICHD is involved with a number of relevant smaller studies, a few interagency efforts (school readiness interventions; developing school readiness outcome measures), as well as a very fascinating study that focuses on rural low-income populations. This study is particularly important because there are not many studies that focus on rural children.

Martha Moorehouse spoke from the perspective of the Office of the Assistant Secretary for Planning and Evaluation (ASPE), HHS. Within the Department of Health and Human Services, there is a lot of inter-agency work. Her office is particularly interested in understanding the specific ways in which quality affects child outcomes. There are a number of groundbreaking studies that will be coming out shortly where quality was measured, then practice was changed, and then child outcome data have been collected. It is important to begin thinking about how to translate what is coming out of these new studies. How can information be moved along for policy and practice purposes in useful ways?

Beth Caron presented on research initiatives within the Department of Education, Office of Special Education Programs. Even though the Department of Education is working on the same issues as the Department of Health and Human Services, their approach is different in that they do not work directly with programs, but

provide federal dollars to the states and fund them. They have put a system into place where states have to set their own targets/goals that are 6 years out and identify ways in which they are going to improve. States must produce annual performance reports for the Department of Education that cover a number of different indicators. The Department of Education then monitors the data that states provide. Whereas in the past, the indicators have been traditionally compliance indicators, they are moving more in the direction of performance indicators.

Inclusion and placement of young children with disabilities is a very important issue for the Department of Education but they do not want this system to be a special education system.

Ivelisse Martinez-Beck briefly shared the research efforts of the Child Care Bureau. The Child Care Bureau is focusing mostly on trying to understand what works to improve quality. They are funding several studies that are looking at this issue (some that are looking at changing global quality and some that are looking at changing quality in specific domains). A new study that will be coming out in a year and a half is looking at what it takes to implement quality enhancements and the cost-effectiveness of doing so.

Discussion. A brief discussion followed the presentations, where the audience had the opportunity to ask questions and raise concerns from their point of view. The major discussion points are outlined below:

- States are still working on generating data and refining their systems; the Department of Education has no intention of ranking states according to the data they provide, but once the data quality becomes good, the Department of Education will work with states to improve their numbers
- No Child Left Behind regulations will get extended to early childhood because people want to know what they are getting for the resources they spend; it is important to be prepared
- The Department of Education and the Department of Health and Human Services need to continue working together

Plenary: Research Perspectives on the Measurement of Quality

The goal of the afternoon plenary was to offer perspectives on the current status of the field with respect to the measurement of quality. Kim Boller moderated the session. Three panelists were asked to discuss the progress of the field in capturing quality across the full range of settings and for all groups of children. They were also asked to address the most pressing needs in the field, the dimensions of quality that are captured well and those that are missing, and promising evidence from new measurement tools.

Barbara Goodson discussed the benefits and limits of the Environmental Rating Scales, and suggested areas of measures development that should supplement the ERS measures. The Environmental Rating Scales' 7 rating scales have become an

institutionalized definition of excellent care and include safe, healthy, and stimulating physical environments; responsive interactions with caregivers; adequate materials; child choice; and to a smaller degree adult needs, individualization, supports for special needs, and inclusion.

It is important to build new information that we have about the importance of the early childhood environment into our established measures. We are now concerned about identifying the *effective practices* that can actually improve outcomes for children, which requires more than mere descriptions of the status of care.

There is currently a great focus on language and literacy in early childhood programs. Even though children need to be exposed to language and literacy, literacy supports in ECE programs are often very limited, even in high quality programs. Furthermore, there need to be supports for English Language Learners (ELL), training in self-regulation, adult responsiveness to children, and supports for math concepts.

In summary, the ECERS does a good job in addressing health, safety, nutrition, and child choice. The ECERS does a moderate job in addressing language interactions, reading aloud, and print exposure. However, the ECERS does not do a good job addressing extensive exposure to vocabulary (an area in which we know low-income and ELL will often be lagging behind their peers), ELL practices, and adult responsiveness.

Measurement beyond ECERS needs to be developed to address:

Emergent literacy
Math concepts
Self-regulation
Responsiveness
ELL practices

Toni Porter discussed the challenges of measuring quality in family, friend and neighbor care—home-based child care that is legally exempt from regulation.. There has been an explosion of interest in this population of child care providers and the quality of care they provide. In response to concerns of state policymakers about health, safety, and support for child development in this type of care, Toni Porter and her colleagues at Bank Street College of Education have created a new measure, the Child Care Assessment Tool for Relatives (CCAT-R, Porter, Rice, Kearns and Mabon, 2005) that addresses quality in these settings. The constructs for this measure were derived from caregivers' own definitions of quality and correspond to many constructs in quality measures that are used to assess quality in regulated family child care and center-based care.

One of the major differences between family, friend and neighbor care and other settings is the nature of an enduring relationship between parent/child and caregiver, which is hard to measure because it goes beyond the physically demonstrative signs of nurturing and sensitivity; it can be a look or the tone of voice used to interact with the

child. Another important aspect of quality in family, friend and neighbor care is shared culture that may differ from dominant mainstream beliefs and practices. These differences may be reflected in caregiver-child interactions related to the use of language and caregiver engagement with the child. This is of particular relevance in light of the growing numbers of ELL children and children from ethnically-diverse backgrounds in family, friend and neighbor child care.

It is essential to assess the care that family, friends and neighbors offer to children, for the very reasons we want to learn about the quality of care in other settings – to understand those areas that may need to be addressed to improve child outcomes and to understand the types of interventions that are effective. However, measures may not capture some of the special dimensions of this kind of care that may have an effect on child outcomes, especially those that relate to social and emotional development for the infants and toddlers who constitute such a large proportion of children in these settings.

Furthermore it is important to consider the "unintended consequences" of not having good measures of quality for family, friend, and neighbor care. If much of the funding for improving child care quality is allocated through Quality Improvement Rating and Enhancement Systems and (other CCDF-funded efforts) that are directed at regulated child care, it is possible that family, friend and neighbor child care will not receive much attention. These kinds of policies may have serious implications for children in those states where family, friend and neighbor care accounts for a significant proportion of subsidized care. These policies may also have unintended consequences for parents who use this care—whether by choice or by constraint—because, by excluding family, friend and neighbor care from rating systems, parents may assume that the care is not good for children.

Peg Burchinal provided an assessment of which quality dimensions are measured well and which not so well by the currently available measures, and the extent to which the current measures are suitable for use across different settings and across different populations. Quality dimensions that are well measured in terms of structural characteristics include class size, ratios, teacher education, economic diversity of the classroom, and time in different settings. In terms of process quality, the well-measured dimensions include warmth/sensitivity/responsiveness of the caregiver, types of activities or centers, and literacy or verbal interactions.

Quality dimensions that are not as well measured include the role of aides/assistant teacher/alternative caregivers in center and family child care, the length of program/use of time, the role of mandated activities on teaching and on measures of process quality, quality of instruction overall and in specific content areas, when curricula can be effective, the role of peers, child engagement, and finally when we should assess the quality of individual child experiences or the quality of overall classroom/setting.

In terms of the extent to which quality measures are appropriate for the full range of settings, Dr. Burchinal felt there was minimal overlap among our quality measures across types of settings and ages of children served. In looking at the appropriateness of measures across different populations (cultural, language, special needs), we need to

distinguish between statistical appropriateness and conceptual appropriateness. From a statistical perspective, the general measures have constructs that cross cultural boundaries and show similar patterns of association with criteria. But from a conceptual standpoint, some absolutely reject that finding and feel measures are needed that are specific to language and cultural norms.

In conclusion, it is unlikely that we will have a single measure that will achieve all goals we have for child care (e.g., sufficient for quality ratings). The measures reflect theoretical perspectives. For example, attachment researchers developed a quality measure that focused on caregiver availability and sensitivity. The environmental rating scales come out of a Piagetian child-centered educational perspective and therefore focus very heavily on children learning from exploring their environment. Now we are seeing a lot more focus on academic skills and so we are seeing more measures focusing on aspects of instructional quality that support academic skills.

We need to learn from the test development industry. They develop a huge pool of items, test the items, and choose according to a set of criteria which items work the best. They refine these items, and test them again. They do this repeatedly until they have a measure with construct validity. We don't tend to spend enough time on item and measure development, or test for construct validity of our items and measures sufficiently in psychology. If we could utilize that type of model, especially for social skills and for quality, we would see stronger associations of measures of quality with child outcomes.

Key points from the plenary papers and open discussion:

- Quality is multidimensional. Warmth may not be sufficient to achieve school readiness, but instead of focusing too heavily on effective instructional practices, we need also to focus on the environment being joyful and inviting.
- Use terms that policymakers will understand. For example, instead of using terms like "responsive caregiving," say "supporting children's development and learning".
- Health and safety are two constructs of quality that cut across settings, age groups, and cultural groups.
- We need to look at both global quality and quality in terms of specific instructional practices (e.g., language and literacy exposure, curriculum around specific topics, etc.).
- Quality is not enough. We have to look at dosage/intensity of the child care experience.
- We need to be concerned about the fidelity of training on quality measures and their faithful implementation first, before we look at changes in outcomes.
- We need to be concerned about giving states measures that they can implement system-wide.
- We're not there yet in terms of measurement development. More work needs to be done on careful item development and testing of construct validity.

- A first step for item development might come from focus groups of different constituencies (e.g., parents, caregivers, administrators, policymakers) to determine what they mean by "quality".
- We need to gain consensus on the constructs of quality that are consistent across setting, age group, culture, and special needs; we need to value cultural differences in terms of learning experiences and perceptions of quality.

Afternoon Breakout Sessions

Four breakout sessions provided the opportunity to have in-depth discussions regarding the current status of the field with respect to the measurement of quality (with an emphasis on the dimensions of quality, quality across settings and for all groups of children, and the sources and cost of data). Each facilitator was asked to bring back to the full group three takeaway messages from each session.

Breakout on Dimensions of Quality. This breakout session was facilitated by Susan Hegland. Bridget Hamre served as discussant. The measurement of quality has expanded beyond traditional measures of global environmental quality to include the in-depth measurement of specific domains, such as supports for early literacy or math, management and infrastructure, and individual children's interactions with the environment and caregivers. This breakout session addressed the question of how these new tools have changed what we know about the important dimensions of quality, and what else is needed in our current system to improve measurement.

What domains of quality are most important to capture in our measurement systems?

- Issues in moving towards measuring more specific domains like supports for early literacy or math:
  - Caution that teachers who do a good job in literacy instruction may not do as good a job at other things.
  - *More time* spent in focusing on specific domains may not necessarily be better for child outcomes. Also look at how well the day is integrated for the child; there has to be some kind of integration. Caregivers need to take into account what children are interested in and build on that.
  - Some of the newer measures of the environment that are domain-specific align more closely with early learning standards and what children are actually being assessed on than some of the older measures that were developed for other purposes.

- What domains were identified for early learning standards across states?
  - A lot of states went with a "back mapping" approach: They started with the K-12 standards and took that backward to preschool and sometimes infants and toddlers. Question whether this is a good practice.

- There are important issues re: how to translate measures to the policy world.

- Measures are not easily "translatable" to the policy world. We need to be explicit about what and how we are measuring and how to most appropriately use measurement tools.
- Policymakers and practitioners want to know what the threshold of quality is for making a difference in child outcomes.

What are the most promising new tools being used to measure aspects of quality?

- The CLASS as a promising measure:
  - Hamre noted that there is clear interest at the state level.
  - Yet systems for training for broad implementation are not there yet as they are for other measures
  - The CLASS is on a 1-7 scale. It has three domains: emotional support, classroom organization, instructional support
  - The average that they see in terms of instructional support is only about a 2; for emotional support a 4
  - There are different uses for the CLASS: regulation, research, professional development.
  - Using the CLASS to guide professional development is easier; using it in regulation is harder. There is no research saying where the threshold is; that is, studies which demonstrate the point at which more positive child outcomes are likely. What we do know at this point is that more is better (higher scores are associated with better outcomes); not how much more is needed to produce positive outcomes.
  - There is some evidence on thresholds based on the ECERS: classrooms have to get up to a 4 in order for things to get better.
  - The threshold issue is important because people think categorically. Cannot separate quality from child outcomes; the threshold needs to be defined based on the child outcomes side. Legislators need to know how much programs need to improve to yield the outcomes they are interested in. But the research base is not there yet.

- Need for measures to capture "the culture of quality"
  - Administrative structures play a big role in creating high quality
  - There has to be support for staff in centers, otherwise staff leave and then classroom quality can't exist. Support and quality are interrelated.
  - Hopefully programs would move toward improving administration if research shows the importance of it. Question of whether administrative support needs to be included as a separate component in QRS.

- Supplementing rather than replacing existing measures
  - Take our existing measures and supplement them with more in-depth measures of quality in specific domains. No need to start from scratch.

- Flexibility of measures to capture dimensions of quality as they are manifested in different settings
    - We have children in very different settings. We need measures that can capture quality in the more informal settings, e.g. after school settings and family day care. We don't necessarily need different measures for different settings: the dimensions in different settings are the same but they play out differently so measures need to take that into account.

- Need to take program resources into account in measuring quality: What aspects of quality take more resources to improve?
    - We need to take the program's resources into consideration. The resource issue is very important when talking about improving low-income settings; they may simply not have the resources to improve some aspects of quality.
    - It is relatively easy to go out and buy 25 books and to count them but it's hard to teach teachers how to use them. It is also harder to measure literacy activities than presence of books.
    - We cannot simply rely on the things that are easy to measure

<u>What are the benefits and challenges of including multiple tools in state quality rating systems?</u>

- The 14 states that now have QRS have each identified a set of key dimensions of quality. States need to continue gathering data and refining their systems in light of the data. Hopefully all states will do that.
- Not all states have QRS but other states have quality initiatives and they also need some guidance.
- The core issue is caregiver/teacher interaction; we need to do the systems work of giving staff the support that they need in order to foster these interactions.
- We should also think about outcomes for parents; what can programs do to engage families and communicate with parents.

The take-away messages from this breakout session were:

- Dimensions of quality are the same across settings but indicators and tools, differ across settings.
- Creating a "culture of quality" – the need for a logic model that specifies inputs and outputs to child outcomes and that links the role of management/ supports to quality.
- Policymakers are going to keep asking about thresholds even though there is limited research clearly supporting them; therefore we are in a position where we need to define thresholds based on child outcomes.
- If we focus only on certain outcomes, we will end up with curricula that only focus on these. We need to be explicit about which outcomes we are really interested in.

Breakout on Settings. This breakout session was facilitated by Cathy Grace. Dawn Ramsburg served as discussant.

Measurement of quality has occurred primarily in licensed or regulated settings and more typically in center-based settings than in home-based settings. This breakout session addressed how measurement tools and strategies can be broadened to ensure that the full range of early care and education and school-age settings are adequately addressed, and whether quality should be defined in the same way across settings. The goal of this session was to develop specific recommendations for improving how we think about quality across settings, considering children's age and characteristics.

<u>What are the most promising new tools being used to measure quality, either across settings or in settings that have received less attention (for example, care by family, friends, and neighbors?</u>

- What we are looking for is good experiences for children across all settings; the question is how we measure these across settings.
- Quality boils down to health/safety, learning activities/instruction, environment, and relationships/interaction. The key issue is how can these be captured across settings?
- Keep parent view in mind.
    - Need to ask parents what they wanted to get them engaged!
    - Tough situation to say that we know better than the parents – role of parent choice Do we ask parents about their satisfaction? If their kids have seemed to grow?
    - Problem: sometimes parents can't believe they are putting their child in bad care. "Parent education" is critical so parents can understand what quality child care is. For some parents, school readiness might not be important - what parents want can be subjective
    - In a study of welfare-to-work, parents wanted affordability and convenience – different conception of what they want/look for.
    - Another example relating to parent beliefs: do we have an expectation that caregiver doesn't do physical discipline? What do we do if parent says it's OK because they do it at home?
- Settings are tied to age – can't talk about quality and setting without taking age into consideration.

<u>Should our expectations and definition of quality differ across settings and in which ways?</u>

- Do we have a clear definition of what settings are?
    - No consensus whether home-based is one thing or multiple things.
    - What about large group homes? With approximately 12 children per home, what does that classify as? Rural states have child care centers that are licensed in the home.

    - Not all states are using the same assessments for the "same" setting, especially when defining settings differently.
    - People sometimes rent an apartment to run a center. Is this a home or center?
    - In multi-state studies, definitions are not the same. Need explicit acknowledgement that each state doesn't define settings in the same way.
    - What about an early childhood program in a school setting? Or a before school or after school program for young children in the school setting?

- Problem with school age settings – few states require care settings in public school systems to be licensed – therefore, if go in and observe, find that the settings don't meet environmental standards.
    - A participant from one state described an RFP process for both private providers and public school providers – have to meet the same guidelines.

- Concern about money spent on measurement of environment vs. on improving it. Issue of resources needed to assure full implementation of program approaches.
    - Does any other field spend more money on measurement than on implementation?
    - Don't measure a program until fully implemented.
    - But provide the resources needed to support full implementation of a program approach.

The take-away messages from this breakout session were:

- There is no one universal definition of settings: e.g. what is considered home-based in one state may be considered something else in another.
- Universal components of quality regardless of setting: environment, relationships/interaction, instruction, and health/safety. What we are looking for is good experiences for children across settings. The question is, how do we measure these universal components across settings?
- Take a serious look at the money spent on measuring quality vs. investing in quality improvements – no other field spends more money on measurement than on creating their service or product. We should not be spending money on measuring the effects of programs if we do not have enough money first to give programs the tools and resources they need to function well.

Breakout on Age, Culture, and Special Needs. This breakout session was facilitated by Gail Upton. Priscilla Little and Richard Fiene served as discussants.

This breakout session addressed the question of whether certain facets of quality are universal, while other facets vary for children differing in terms of age, cultural background and whether they have special needs. The goal of this breakout was to develop specific recommendations for improving how we think about assessing quality across the age range and according to other characteristics of the child.

What are the most promising new tools being used to measure quality across age groups?

- Fiene started the discussion by focusing on the need to consider linkages of structural and process aspects of quality.
    - How do the structural and process aspects of quality fit together, and how does that fit with age, culture, and special needs?
    - As we started to talk about program quality earlier in this roundtable meeting, we could not leave out talking about health and safety. The structural and process aspects of quality need to fit together. Many of the structural aspects fall into the licensing (health and safety) arena.
- Looking at process, the interactions among children and adults in classrooms is central.
    - Fiene and his colleagues have developed a new measure of interactions based on the Arnett scale called the CCIS (Caregiver Child Interaction Scale). This scale is better at discriminating between those who are mediocre and those who are doing a good job (rather than just good and bad, period).
    - The Arnett is a 4-point scale that measures frequency. Fiene's group has now switched it to a 7-point Likert scale.
    - The new measure is going through major pilot testing right now. They have designed an infant-toddler and preschool age format. They will eventually do a school age version. But they are focused on an infant-toddler measure of interactions now because that is where the need is.
    - The revised Arnett measure will be ready by summer of 2007. They switched all the indicators to the positive. Reliability is really high, comparing it to other scales (old Arnett, ECERS, ITERS).
    - Too difficult to get time sampling on this.

- Issue of which tool do we want to use for what purpose?
    - Our current assessments aren't covering some of the constructs or settings that we are concerned with now.
    - Issue of using existing tools vs. developing new tools: Developing a new tool is so expensive.
    - If many studies have used the ECERS, then there is reason to use it for comparison's sake. But the field should have some room for openness, especially for training, self-assessment and PD plans.
    - If you are going to have a state policy based on an assessment, you need a reliable tool. But if you have an individual program that wants an evaluation tool, by all means pick something that is appropriate for that program, rather than be locked into one instrument that has had a great deal of testing done on it.
    - The quality measures don't do a good job of measuring age, culture and special needs.

- What are the common features of quality that cut across the different ages, different settings?

- In one state, work is going on to help the state legislature identify the outcomes for state after-school programs.
  - There is a set of "anchor outcomes" that all programs have agreed to collect (e.g., homework completion, school attendance).
  - There is also a set of "choice outcomes" specific to particular programs (arts, etc.).
- Following this approach, what are some "anchor outcomes" we could nominate that cut across ECE settings?

- When and how are ERS scales used in different state QRS systems?
  - In discussion it was noted that different states are identifying ways to use the ERS scales only at certain levels or at certain time intervals. As examples:
    - In one state the ERS scales are carried out only at the higher levels of the QRS. At lower levels, the provider does a self-assessment.
    - In another state it was noted that because of cost, they cut back from carrying out the ERS every year and are now observing less often.
    - Yet another state does not use the ERS ratings in the determination of points in the QRS, but for self-assessment. Just getting the results/profile report is an intervention.

<u>How can we include measures of quality for children with special needs?</u>

- We aren't really addressing kids with special needs. We kick kids out of child care because they have a diabetic pump.
- We have very little evidence on dosage at all. Some kids are in special education 1 or 2 hours a week. We're going to aggregate across kids and say that in general the program is showing evidence.

<u>How can we ensure that issues of language and culture are captured in our measures?</u>

- Sometimes, we start with the top-down discussions. When talking about culture, we need to start bottom-up and talk face to face with different groups about what they think about child care quality.
- We need to be much more sensitive in what caring for children means in various communities.
- We need more research on the communities we are serving.

The take-away messages from this breakout session were:

- Age, culture, and special needs need to be part of the discussion when talking about quality and not just looked at in isolation
- Different cultures view quality in different ways; in order to take into consideration how different communities define quality, we need to talk to stakeholders in these communities

- Standards of quality are relevant across age groups, settings and cultures, but indicators of quality may be different in these
- A decision tree is needed for researchers, policymakers, and practitioners to know which measure to use for which purpose/setting/age group

Breakout on Sources and Costs of Data. This breakout session was facilitated by Kathy Thornburg. Helen Raikes served as discussant.

The measurement of quality in a policy or program context must strike a balance between the use of reliable, in-depth tools with the need to use measures that are cost-effective and feasible in terms of the time and training needed to administer them. This breakout session addressed what research can tell us about where states should allocate their scarce resources.

<u>What do we need to know about quality to inform out understanding of child outcomes?</u>

- Need an observation somewhere in the system.
- Need to consider how quality rating system is related to licensing system.
- Why haven't states responded by only using one or two measures from the ECERS instrument to cut down costs? Why use the whole thing?
    - Some research shows you can pick 17 items, but there may be sampling issues
    - The other issue is what message are you sending by what items you choose to observe.
    - Also, you need a 3-4 hour observation period anyway, so why not use the entire instrument.

<u>What is the threshold at which the prediction of child outcomes is no longer boosted by the addition of new or more in-depth data?</u>

- Observation certainly involves high cost. But there are also costs involved in verifying report data.
- Providers are frustrated that results take so long to be produced, but having one small group of raters can help speed up analysis and reporting
- In some states there are registries that provide on-line records of professional development. Linking with registry data means that raters do not have to spend the time and cost doing that data entry themselves
- It works to implement an online system for licensing as well.
- Having linked data and being able to submit queries across administrative data systems would be beneficial, but difficult; ideally they would like to have a single database system

<u>What innovative measurement strategies are yielding promising results?</u>

- The ECERS-E seems promising, as well as early learning guidelines.
- There is a concern that the ECERS is becoming our definition of quality, something that may go beyond its original intent.

- There are different strategies for sampling classrooms:
    - Some states average scores across classrooms but make a requirement that no classrooms can get below a certain score.
    - One participant noted that in her state they had data from every classroom. In order to examine whether taking a sample of classrooms would provide a reasonable picture, they did an average based on all classrooms, based on one third and based on one half of classrooms. Average scores were very close. Based on these findings, the decision was made to sample.
    - When sampling, however, the problem may arise that classrooms not observed nevertheless want feedback.
    - In some states, a difference in reliability of those conducting ECERS for TA vs. for research has become an issue. As a result, several states now train TA staff in the same way as monitoring staff to assure that feedback to providers is in alignment with scores completed for monitoring. It has worked well for quality improvement to train TA to do the ECERS.

The take-away messages from this breakout session were:

- There needs to be more discussion on using the same tool for multiple purposes (research, quality improvements, etc.). Issues can arise when different ratings are reached by observers trained to different standards of reliability. Consistent information on program quality is important.
- No one tool should become the definition of high quality
- There are potential strategies for saving costs on data collection and analysis so that more dollars can go into improving programs. One possibility is data infrastructure – linking databases to cut costs. Another possibility is sampling classrooms to save on observation costs.
- There is a need for a body of knowledge on best practices in data collection within quality initiatives. Issues can be looked at empirically and the information shared. For example, how many classrooms need to be sampled within a center to obtain an overall picture of quality? Similarly, we need empirical examination of how reliable self-report data are, and whether periodic auditing of self-report data would affect reliability.

WEDNESDAY, DECEMBER 6, 2006

Morning Breakout Sessions

Four breakout sessions provided the opportunity to have in-depth discussions regarding key outcomes: the effects of quality initiatives on programs, providers, parents, the market, overall quality, and children's school readiness. Each facilitator was asked to bring back to the full group three "takeaway messages" from the breakout session. We summarize the key points from each session.

Breakout on Programs, Providers and Quality of Care. The facilitator for this discussion was Kathy Modigliani and the discussant was Sharon Ramey. The goal of this breakout session was to identify key research questions about the effects of quality initiatives on programs, providers and the overall quality of care within a state or locality. The questions that were discussed by the group and the key themes/issues that were raised in response to each question are detailed below.

Which aspects of quality are most difficult for programs to improve over time?

The group identified the following aspects of quality as most challenging to improve: Interactions with children to promote language and learning, interactions with children to promote behavior management (though with intensive intervention/training, changes can be made), providers' openness/readiness to change, implementing appropriate systems to run a high quality program (record keeping, administrative and fiscal systems, supervisory systems, systems to take advantage of community resources), observation of individual children and using observations to adapt/create appropriate programs, and provisions for children with disabilities (or even just challenging behaviors).

Are specific groups of providers more and less likely to participate in quality initiatives?

The group concurred that providers are more likely to participate in initiatives when they have higher education levels and already attend training. Providers less likely to participate may not have a financial incentive (e.g., they do not/will not serve subsidized children). Other providers may not be interested in licensing or star ratings. Some rural providers may not be motivated to improve because the supply of providers is not large enough to offer a wide range of choices to parents. Other reasons providers may not participate include language barriers, low educational qualifications (for example, those who did not do well in school may be less likely to take classes or training), cultural differences in childrearing/educational philosophy, a fear or anger about not being designated at the "top level", the expense associated with the quality indicators (for example, offering a pension program for staff), and providers/programs, such as Head Start, who may not consider themselves part of the "child care" system.

The group considered other issues related to participation in quality initiatives including the importance of determining that providers are translating new knowledge

into their practice, the need to include recent immigrants/non-English speaking providers in initiatives and addressing turnover among home-based providers.

What strategies work best to engage providers of all types in quality initiatives?

The group discussed the importance of two general types of approaches. One involves on-site, intensive coaching (including demonstration of interactions with children) that is coupled with provision of resources and focusing on child-adult interactions and health and safety issues. A second approach involves the use of peer mentors or peer support in a variety of ways including accreditation facilitation, group discussion about what works and what doesn't and networks of peers. Both approaches need to be implemented using research on adult learning principles.

What are the unintended consequences of quality initiatives on programs and providers (e.g. shifts by providers out of licensed care)?

The group discussed the potential for a number of unintended consequences including: exit of providers, particularly license-exempt providers, from the subsidy system; exit of more highly educated providers from child care because they qualify for higher-paying jobs; movement of enrollment from centers to family care providers with fewer quality requirements; potential "fast tracking" of providers through training without any demonstration that they are implementing new skills; and; increased certification requirements for training may push people who would be/had been great trainers out of the training system.

Breakout on Markets, Access, and Cost. This breakout session was facilitated by Helen Raikes, with Michelle West serving as a discussant. The goal of the session was to identify the key research questions about the effects of quality initiatives on the supply and demand, cost, and access to early care and education and school-age programs. From the perspective of states and communities funding quality initiatives, what are the most important market outcomes to examine?

The group focused their discussion on the following questions: From the perspective of states and communities funding quality initiatives, what are the most important market outcomes to examine? What research designs would help determine whether "scholarships" or increased subsidies are needed to ensure access to higher quality for low-income families? What research designs examine whether the balance of licensed and legally unlicensed care utilization shifts for families receiving subsidies with the introduction of quality initiatives? What research is needed to provide estimates of the cost of improving the quality of care across different levels? What research would help states make appropriate estimates of the cost of outreach and ongoing monitoring of quality initiatives? How can research be structured to ensure that we are measuring the unintended consequences of quality initiatives on markets, access, and cost (e.g., increases in cost as demand for high-quality increases, decreased access to high-quality care in low-income communities)?

The following is a summary of the key points (and potential research questions) raised in the discussion.

What are the most important market outcomes to examine?

The group discussed parents' choice/use of care as one important outcome to examine. Key questions to examine around parents' selection of care included: how do parents use quality information? Which pieces of information matter most to parents? How do choices and use of information vary by parent characteristics such as education, culture, income? What is the elasticity of child care costs (how much more are parents willing to pay for high quality)? The group also highlighted a number of research design issues including the need to sample both parents who used information available in a QRS and those who have not. The group also discussed a suggestion to track a group of parents who chose a high quality setting and to ask them about their decision-making process. Focus groups may work best for this type of question.

Program participation was another outcome discussed by the group. Key questions to examine included: which programs choose to participate in quality initiatives? What is the financial gain for programs? Do they have enough children receiving tiered reimbursement to make a financial difference for themselves? The group also discussed the important of understanding more about the geographical characteristics of programs that do and do not participate in quality initiatives. Where are high quality programs located? How convenient are the locations for low-income families who use public transportation? What can we learn by examining "saturation" or "penetration" of quality into a geographical area? Geocoding was proposed as a method for examining some of these questions.

What mechanisms are most successful in sustaining quality?

The group raised the importance of evaluating the success of strategies such as tiered reimbursement and public-private partnerships in helping sustain quality improvements.

How can research be structured to ensure that we are measuring unintended consequences of quality initiatives on markets, access, and cost (e.g. increases in cost as demand for high quality increases, decreased access to high quality care in low income communities)?

The group discussed a need for a cross-state study that can examine variations in key features of quality rating systems. In some cases, variations can be examined within state as well.

What research designs would help determine whether "scholarships" or increased subsidies are needed to assure access to higher quality for low income families?

The group discussed the importance of understanding the combination of incentives and scholarships that might work best. It may be important to experiment with different combinations or packages.

Are more families going into licensed care? Are there unintended consequences?

The group discussed the need to use longitudinal methods to track parents and children with subsidies. It will also be important to track patterns of care usage to know where children with subsidies are going, and if possible, to identify patterns by the quality designation of settings.

What research is needed to provide estimates of the cost of improving the quality of care across different levels?

The group concurred that it would be best to examine these costs as part of a study involving multiple states and to understand costs at different levels of analysis (for example, individual facilities' costs, and professional development systems' costs).

The group concluded this breakout session by discussing studies that they would recommend.

> The first study would be a collaboration across multiple states. The intent would be to address similar questions and use similar methodological approaches in differing contexts in which QRS are being implemented. Questions would include *change over time* as QRS are fully implemented, in terms of: (1) the distribution within a state of care of different quality levels, (2) access of low-income children to care with higher quality ratings, and (3) participation of subsidy-receiving children in care with higher quality ratings.
>
> The second study would be a longitudinal study of parents, focusing on whether parents' understanding of the components of quality change with the implementation of a QRS, and whether and how the quality ratings are used in making decisions about care. It would be important to examine this question for parents with children of different ages, and in states varying as to resources allocated to outreach to parents in different income and cultural groups to inform that about the QRS.

The group decided that a consortium of studies would be needed to capture the naturally occurring variations and to examine implementation questions.

Key themes from the discussion:

- It is important to come to agreement on a set of measures to track regarding how QRS are functioning within the child care market. Potential measures that QRS are functioning as intended identified in the breakout group included: parents are selecting higher quality programs; low income children have better access to high

quality programs; overall number/proportion of programs rated at higher quality levels goes up over time; more programs are becoming part of the QRS.

- The group recommended studies to capture costs. Key questions include: what are programs spending to participate in QRS; how much money is being spent by states on incentives, technical assistance, measurement of quality. Key questions also include whether the price for participation in programs change and whether provider pay changes over time with the implementation of QRS?
- The group proposed a consortium of state studies to look at variations in QRS and how these map against the measures of QRS functioning.

Breakout on Parents. The facilitator was Deborah Ceglowski and the discussant was Bobbie Weber.

The goal of this breakout group was to identify the key research questions about parents' reactions to quality rating systems and other quality initiatives. The questions to be discussed were as follows: What are the key outcomes for parents that need to be addressed by research? How can research address the key question of whether parents find information about quality useful, and whether and how they actually use the information in choosing care? What research approaches would be needed to identify whether there are subgroups of parents more and less likely to utilize quality summary information? Can research help to provide information on the most effective outreach strategies for informing parents about quality initiatives?

Deborah Ceglowski's suggested that the group use Lillian Katz's model to frame the discussion. Looking at quality from a research perspective could be considered a "top down" perspective. Using Katz's model, alternative perspectives on child care quality can be identified, including the "inside out" perspective (children's actual experiences in child care), the "outside in" perspective (care from the perspective of staff or caregivers) and a "bottom up" perspective (parents' perceptions).

The group discussed the "bottom up" perspective (parents' perceptions) by posing questions about: the ways in which QRS information and data will be used by parents; the ways parents will combine information from different sources; the weighting that parents assign to information; parents' perceptions of the information content that will be most helpful, as well as the format and structure of the information.

The group discussed the importance of examining nuances in how parents' perceptions of child care may change over time depending on, for example, the age of their child and their previous experiences with care. It will be important to study these variations. The group also felt that information about quality needs to be specific (rather than a "general picture") using data that allows comparisons to be made between programs.

In the discussion on which subgroups of families are more likely to access quality information, the group shared perspectives from resource and referral counselors in their states as well as from providers on the constraints and choices parents have when

selecting care. In their questions to resource and referral counselors, some lower-income families need to prioritize proximity of a child care setting to their home or work over other factors. Parents are also concerned about keeping siblings in the same program. Parents using child care subsidies also worry about the availability of slots in high quality settings because some providers are reluctant to engage with the subsidy system.

The group then posed additional questions about parents' perceptions of a quality rating. For example, how does it feel to parents to find a place for your child and find out it's a "low quality" program? How do they use the information, and what do they do when other issues (cost, accessibility) need to override the information on quality?

The group discussed the option of research that examines how parent involvement in a program can affect quality. The group questioned current strategies for measuring parent involvement and wondered if there is evidence about the validity of these measures.

The group proposed the notion of assessing not only construct validity but "parent validity"…how valid are these measures in the eyes of parents? The group noted that Art Emlen has developed scales to assess quality from a parent's perspective.

There are a number of interesting issues that need to be considered in the context of "parent validity". It was suggested that parents ultimately have a perspective on a child care setting that is influenced most by their perceptions of how well their own child is doing in the program and their own experiences in communication with the program or provider. Important constructs such as the power differential between the provider and the parent may need to be considered as further work on parents' perceptions moves forward.

A final point was made to distinguish the terms "measurement" and "documentation". In most quality rating systems, aren't we using documentation rather than measurement?

As a model of innovation, the group discussed an "indicators project" in a state that focuses on providing information to parents but not in a quality rating system.

The group summarized the key themes of the discussion as:

- The importance of looking at "construct validity" from the perspective of parents.
- The need to frame parent choice as parent use (which may be a more accurate term)
- The centrality of the relationship between parents and providers/programs and the importance of accurately assessing parent-provider communication

Breakout on Children's School Readiness. The group was facilitated by Jean Layzer. Barbara Goodson and Peg Burchinal served as discussants.

The goal of the breakout session was to identify the key research questions about effects of quality initiatives on children's school readiness. The questions suggested for discussion included: What are the most important child outcomes to examine? What assumptions about duration of exposure to different quality levels need to be met in research examining the linkages of quality initiatives and children's school readiness? What research approaches are feasible for examining QRS and other quality initiatives and children's school readiness? For example, are there possibilities for comparative designs?

Jean Layzer began the session by saying that states are most interested in the child outcomes that can be achieved by their quality initiatives. She invited the group to discuss current research and new research that is needed. She then introduced Barbara Goodson and Peg Burchinal and asked them to discuss research in which they are involved.

Barbara Goodson provided details about three projects she is working on that represent specific interventions focused on specific outcomes. Findings from the three studies are not yet available publicly but, when they are available, they will provide important information about the degree to which child outcomes can be affected by interventions.

Peg Burchinal spoke about the QUINCE evaluation (with results not yet available) and about the NICHD Study of Early Care and Youth Development. From the observational studies conducted as part of the NICHD study, Burchinal described analyses to look at what aspects/measures of children's development seem to be related to child care settings and are in turn linked to success at school. Using path analyses, one paper examined whether and how child care quality experiences were related to children's reading and math scores in 1st and 3rd grade. The paper looked at language, math and continuous performance task (CPT) – a measure of self-regulation. Reading appeared to be the most important mediator. The significant path was from child care quality (as measured by the ORCE) through reading to children's later reading and math skills. There is less evidence that quality was related to self-regulation.

Another study looked at predictors of academic success at grade 5 (trajectories from 54 months through grade 5). The strongest predictors were language and self-regulation. Social skills were important, but not as important. Burchinal described the potential for measurement problems playing a role in the strength of the findings because the measures of social skills were indirect measures (teacher report- the California Social Competence Scale), whereas the language and self-regulation (measured by the CPT) were measured directly by child observation/assessment.

Jean Layzer then invited the group to engage in discussion. The following bullets summarize the key themes that emerged in the discussion.

- The measurement issues around testing the effects of quality improvement interventions on children are complex. The measurement is often concurrent, with quality and child outcomes being "moving targets". Effects are often examined by looking at how changes in the learning environment or caregiving practices are related to changes in child outcomes. It may be important to look at child outcomes after the caregiver has been trained and do the post-test at a later time. For states, this means that time will need to be allowed for assessment of children.
- Family factors are extremely important to consider in the assessment of child outcomes. These will account for more variance than child care experiences. Yet child care experiences are easier to change than family characteristics. An approach to consider involves examining the effects of a program for different subgroups of children (for example children who are low-income but not in the social service system; children who are low-income and are in the social service system; and, children who are not low-income). It will be important to understand what the baseline is for different groups of children. Collecting data from programs on percent of children who are English Language Learners or who have IEPs will help to understand the baselines.
- Selection bias is a very serious problem in our studies. A recent meta-analysis conducted by Abt Associates showed that effect sizes are bigger when studies are less controlled. States are not always in a position to use random assignment studies, but it will still be important to think about the data that can be collected to estimate selection bias.
- It will be important to convince state policymakers that the collection of appropriate data on child outcomes will require time to collect. They also need to have reasonable/realistic expectations about the changes in child outcomes that can be expected from improvements in child care quality.
- Cost effective processes are needed for assessing children. Using teacher report is problematic because of bias. Options for using teachers as assessors include the provision of training and random checks, but the qualifications of teachers need to be considered. For example, these processes may be more successful with BA level teachers. The National Reporting System conducts a quality assurance study and finds very high reliability in the assessors, though the assessment is very scripted (leaving very little room for error).
- It is important for states to clearly define the questions on which they need answers because this will determine the measurement strategy. For example, understanding how investments impact children requires an approach that is different from asking if quality affects children's development.
- New data collections are costly – states need help with strategies to define narrowly how much original data collection needs to be done and where to look for data that is already collected
- There was a recommendation to conduct small randomized studies to test what you want to do. We were successful in persuading Miami-Dade to wait and do a test of several variations on the strategy to see what worked best in their setting. The results will justify waiting; it certainly saved them a lot of money.

- States making investments in QRS need to balance their investments across all related areas (e.g. program assessments, technical assistance, training, etc.)

The group then turned to a discussion of the take-away messages from the discussion. The points that were raised are summarized below:

- Investment in quality enhancement needs to have sufficient time to be implemented before we expect to see outcomes change. Policymakers need to know the sequence of questions that have to be answered first before child outcomes can be examined.
- It is important to collect information on family background as well as on the type of program children are in.
- A research agenda should be laid out with a schedule for time and cost. It needs to be stated clearly without research jargon. Intentions and expectations also need to be stated clearly without being unrealistic or overselling the potential effects on children.

Town Meeting: Reflections on Key Themes and Issues

Lori Connors-Tadros, the moderator, introduced the session by saying that the meeting has created a space for dialogue among researchers, policymakers and practitioners. A goal is to take the information back to the individual work meeting participants are doing and infuse the lessons learned. Connors-Tadros highlighted the importance of meeting conversation about the definition of quality and how quality is linked to outcomes. She then introduced the four panelists who were each given 5 minutes to discuss their reflections on the meeting and how they will be bringing information back to the ongoing work of their organizations.

Jerlean Daniel, NAEYC. Jerlean Daniel described NAEYC's accreditation system as, in essence, a national quality rating system that's purpose is quality improvement. When accreditation began 20 years ago, NAEYC was told that it would be too hard to encourage programs to engage. Yet, Daniel described surprises in their history of accreditation in terms of engagement in the system and told the group not to underestimate the pull of a quality rating system.

NAEYC accreditation has been newly reinvented, and she asked the group to weigh in on any domains or indicators that need to be reconsidered. In the reinvented accreditation, the issue of culture has been included throughout the ten program standards. In addition, inter-rater reliability requirements have been added.

Daniel noted that NAEYC is aware of the issues raised in the meeting and has a number of initiatives that are aligned with the issues that have been raised. For example, NAEYC has standards for teacher preparation and now has an accreditation process for Early Childhood Associate Degree programs. They are offering scholarships to help

offset the cost of accreditation. They would like to see NAEYC accreditation not just as a part of quality rating systems but at the top of these scales.

Daniel anticipates that the new system will have a large database of information about programs that will also be useful to the research field. She also announced a new applied research initiative at NAEYC with a staff person who will be responsible for helping NAYEC use their own database more efficiently.

Ellen Frede, National Institute of Early Education Research (NIEER). Ellen Frede described her perspective as a "state pre-K lens" because it was the focus of her previous work and is her current focus at NIEER. She described the focus on pre-K as in fact a "blurred distinction" between pre-K, child care and Head Start (since, for example, many state pre-K programs are offered through child care centers). Pre-K is intended to be educational, though quality may actually be lower than quality for child care.

Frede noted that an issue that arose at the meeting was unintended consequences. She highlighted the need to think more about the effects of pre-K initiatives on child care availability and cost. A second theme she heard was the need for better data and better data infrastructures. She also noted the dearth of good outcomes measures and described a new measure she has developed to look at math (the Preschool Classroom Mathematics Inventory).

Frede said that NIEER would like to capture more of what is happening in child care with respect to state pre-k and that an important step may be to bring together state administrators to discuss child care/pre-k connections and the research agenda that should be developed.

Marlo Nash, United Way. Marlo Nash described her work overseeing two national initiatives at United Way. One is a community impact model in which communities identified quality improvement systems and accreditation projects as an essential component of developing quality. Local communities were realizing that low performing programs needed "stair steps" to help move them toward quality, so United Way is helping to take leadership on this issue at the national level

Nash described the United Way as being at this junction of quality research and practice. They are convening business and community leaders with practitioners and researchers. Business leaders are excited about the changes that can be made in early childhood programs but they want to see the evidence of improved child outcomes (not program outcomes). Nash described a sense of weariness with descriptions of investments in children that aren't accompanied by description of how their lives have changed.

Nash described an opportunity to get business leaders involved in influencing legislative leaders. She described the need to be able to articulate a vision at a high level so that, for example, even a 5-minute discussion with business and legislative leaders will motivate interest and excitement.

Jana Martella, National Association of Early Childhood Specialists in State Departments of Education. Jana Martella has worked for five years on early childhood initiatives, and more recently with early childhood specialists in state departments of education. Martella described feeling very welcome at the meeting and felt that the information discussed was extremely meaningful for her audience. Martella made a plea for developing information that can be meaningful for practitioners.

Martella agreed with Marlo Nash that there is a sense of urgency and momentum around the issues discussed at the meeting. She also senses fatigue and a feeling of needing to do more with less. She discussed the need to recognize the resources (including human capital) that are needed to address the issues.

Debbie Moore, National Association of Family Child Care. Debbie Moore described her primary concern as getting family child care providers recognized as a central component of the early childhood education system. She described as particularly salient to her the discussion of who family child care providers are and where they are placed on the spectrum of home-based care. NAFCC is working to improve quality across the whole spectrum. Moore said that she is looking for more research on how to recruit, support, and retain intentional providers. She also raised the issue of collective bargaining and what the outcomes of the system are going to be across the board. She described the importance of tying increased wages and benefits for providers to increased quality of care for children. Moore described findings from Maryland kindergartners that show similarities in outcomes for children coming from family child care and children from center-based programs. She concluded by suggesting that legislators may want to reach a threshold of "good enough" as opposed to best practices, and the field should work to find a balance on this issue.

Questions and Comments from the Group: Connors-Tadros invited questions and comments from participants at the roundtable. The following bullets summarize key themes that were discussed.

- It will be important to convene a group of experts to assess which aspects of quality are most important to invest in. The field needs to consider the cost and anticipated benefits of certain quality investments. These decisions should be based on our best summary of the available evidence and confidence that investments will matter for child outcomes.
- A subsequent suggestion was offered recommending the model of a "consensus conference" (such as those convened by NIH) in which unbiased participants with a mix of real world and scientific expertise conduct a research review. The conference could consider the facets of quality that are most important for children's outcomes. The consensus group would examine the evidence to support certain quality indicators and report on whether an indicator is known to produce positive child outcomes, whether the evidence on the indicator is mixed, and whether the indicator does not appear to be contributing to child outcomes.

- Regarding measurement of child outcomes, a question was raised about natural developmental differences in children, especially in the preschool years. Children develop at widely differing rates and learn in very different ways. By focusing on outcomes, are we raising unrealistic expectations for children? The group discussed the need to rely on direct measures of skills after age 3 when the curve of individual differences becomes more normal. We should be concerned about assessments that rely only on teacher ratings because of the potential bias in these measures.
- The group discussed the importance of examining family background characteristics and selection factors when studying children's outcomes. In addition, it is necessary to recognize starting levels, that is, where children are upon entering programs and what progress is reasonable to expect.
- A range of research and evaluation strategies are used to assess how quality affects children's development, and our confidence in findings should be based in part on the strength of the research design.

Moving Forward: How to Create and Maintain a Process of Ongoing Dialogue

Kathryn Tout served as the moderator for this session, and Bobbie Weber and Marty Zaslow served as panelists. Tout introduced the panelists and thanked the meeting participants for their contributions to the meeting. The panelists were asked to provide a synthesis of key themes discussed at the meeting and their perceptions about opportunities for maintaining ongoing dialogue among meeting participants.

Bobbie Weber. Bobbie Weber began by saying that it was difficult to summarize the meeting proceedings but that she had a number of reflections on the conversations she heard during the meeting. She felt consensus in the group that quality improvements are needed for all children, particularly for low-income children who may be more affected by the quality of care they receive.

Weber raised some conceptual and linguistic issues that arise with the term "quality". The term evolved to describe care and education that is associated with good child outcomes. However, over time, we have realized that there are complexities with this term. There are multiple domains of quality and multiple child outcomes that are of interest. Can a single term communicate these complexities? Weber heard a concern expressed at the meeting that one measure of facility quality should not inadvertently become our definition of quality, and further efforts need to be made to develop better terms that convey the complexity of the concept. She reminded the group that rating systems are only one component of quality improvement efforts. They are not a stand-alone effort, and they will not have efficacy by themselves.

Weber conveyed the centrality of the providers as the drivers of change for children – quality investments can be made by parents, providers, funders and community organizations, but providers themselves need to be the focus. Weber stated that we don't know if the provision of information or measurement itself can actually change providers' behaviors.

Weber highlighted the need to consider multiple types of care (particularly care by family, friends, and neighbors), the full range of children's ages (for example, the unique circumstances of school age children), and the importance of understanding how cultural and language differences can be incorporated into measurement strategies.

Weber raised the issue of evaluation that can disentangle the unique effect of investments versus the provision of new information. Which is more important? In addition, little empirical work has been done to assess whether quality initiatives are making a difference in the decisions of parents, providers and policymakers. Which parents are using the information? How are they using the information? How does it fit into the complex set of factors that go into choosing child care?

Weber also stressed the need to consider the multiple components of state systems that play a role in quality and the importance of clarifying roles and relationships. She gave the example of licensing and quality standards and the need to avoid confusing messages and duplication of efforts. There was a consensus that health and safety is a foundation of quality, but there was no apparent consensus on whether these aspects should be included in licensing, quality standards or both. Weber also asked for evaluations that consider the effects of quality initiatives on low-income children and the settings they use.

Weber concluded her comments with a discussion of states' needs and the guidance they are looking for on how to make investments in quality. She described the following options: Create a list of quality investment options; convene a group to articulate a set of principles or guidelines for states that would help them create a balanced portfolio of investments; identify which facilities and children will be reached under different scenarios; create a compendium of facility assessment tools, perhaps using an interactive, web-based platform that includes a "decision tree" (to guide decisions), true costs (in terms of human capital and other resources), and reliability/validity information; and, compile a similar compendium for assessment of child outcomes that is nuanced but still simple and clear for use by state administrators.

Marty Zaslow. Marty Zaslow framed her comments by describing two metaphors that had become salient to her during the meeting: a ladder and an escalator. Zaslow described each of these metaphors in turn.

The ladder metaphor can be used to represent the measurement of quality. During the meeting, questions about how the rungs are defined and the location of the bottom and top rung were articulated. Across several presentations, a picture emerged in which the ladder seemed to have three different segments. Measures of health and safety are seen as foundational: the bottom segment of the ladder. Whether they are measured through licensing or through a quality rating system, there was consensus at the meeting that it is critical that they be measured. Interestingly, measures of health and safety may not survive if quality measures are chosen on the basis of how individual indicators of quality group together (e.g., through factor analysis): they do not tend to correlate with

other facets of quality and tend to be dropped when measures are chosen based on empirical groupings. Another key issue noted at the meeting is that we often fail to look at child health outcomes in studies linking quality and children's development, so we may also miss these measures if we use prediction to child outcome as the criteria. Nevertheless, there was clear agreement that these are central.

Regarding the middle and upper segments of the ladder, Zaslow noted that several presentations at the meeting portrayed our traditional measures of quality, such as the environmental rating scales, as the middle set of rungs on the ladder. They are beginning to be seen as needing to be complemented with a further set of measures. These measures that Zaslow referred to as "reach" measures following up on discussions at the meeting, attempt to capture intentional instruction in early childhood settings, such as stimulation for early literacy or early math development. Measures here are still in development.

The presentations at the meeting acknowledged that our existing measures of quality are showing small or moderate associations with child outcomes. The discussions at the meeting raised the possibility that the more specific "reach" measures may be more closely aligned with the facets of children's development we are measuring as outcomes. The extension of the ladder to capture a fuller range, and measures of quality that are more closely aligned with the child outcomes we are measuring, may result in stronger associations between quality and child outcomes. This is a possibility that should be followed as research now in process is reported out. The meeting presentations underscored the importance of a new set of experimental studies now being reported out or shortly to be reported out. These studies will help ascertain whether different approaches to improving quality indeed to do so, and whether they also result in improvements in specific facets of children's development. These studies will also provide a new perspective on the strength of the association of quality and child outcomes.

Zaslow noted that some of the issues discussed at the meeting had to do with how best to go about collecting the data for the different rungs of the ladder as these are defined in different states and communities. We are at the point where we can begin to learn from careful examination of such issues as how many classrooms need to be observed in a child care center in order to obtain a portrayal of overall quality; how often interrater reliability needs to be obtained; how often on-site observations of quality need to be carried out; and whether and how self-report data used in quality rating systems need to be verified. Other important issues about how measures of quality are collected concerned making clear distinctions between how quality measures are collected for monitoring and accountability purposes and for purposes of providing technical support to improve quality. As states begin to collect data on quality on an ongoing basis, we need to summarize what is being learned about such issues through empirical examination of them and as best practices emerge.

Zaslow then introduced the escalator metaphor. The participants at the meeting repeatedly returned to the theme of not just measuring or rating quality effectively and efficiently, but also understanding what results in upward movement; thus not just a

ladder but systems that were more like escalators. She noted the fundamental question of whether just the provision of better information on quality would result in providers opting to improve quality, parents choosing higher quality, and changes in the market. We need to know whether it is essential not just to provide information on quality, but also a set of incentives, supports and requirements. We heard about different state approaches to providing these, with explicit discussion, for example, of incentives and requirements to assure that quality improvements reached subsidy-receiving children. With the emergence of different state approaches to fostering the upward movement of quality (escalator and not just ladder), it becomes particularly important to follow up on the present meeting with ways to learn about how state variations function. We need to consider ways to continue discussions so that we can look across the experiences and research results in different states, to learn what program variations are most important to supporting improvements in quality and access of low income children to the higher levels of quality.

Zaslow noted the emphasis that was placed during the meeting on assuring that efforts to improve quality were structured so that they reached care settings serving low income and subsidy receiving families. She also noted repeated discussions about how the concept of an escalator could be applied to family, friend and neighbor care. What approaches to engaging these caregivers will be effective in improving quality? A first set of evaluation studies looking at attempts to improve the quality of family, friend and neighbor care should be informative.

Zaslow concluded by saying that we will be learning by doing. We need mechanisms to continue to exchange information on the experiences of states and communities and on research focusing on some of the key issues that emerged at the meeting. Research Connections will be a key resource on the emerging research, and we can use the National Child Care Information Center as a critical source of information on what is happening in states. Participants at the meeting have raised the possibility of systematic strategies to review and synthesize emerging research findings, and this possibility should be given serious consideration.

Group discussion

Kathryn Tout opened the discussion to the full group. A number of final comments were offered by the meeting participants. Key themes from the comments are summarized below:

- Some of the questions posed as the meeting may be answered with relatively small investments (for example, secondary data analysis on how QRS has affected participation, uptake, etc).
- It will be importance to share information within and between states. From this information, it will be important to develop information on best practices in data collection as well as structuring of quality initiatives.
- Recommendations should be based on a body of work, not on single findings. We need to be clear about where the research is limited.

- An independent, unbiased research group may be needed to convey the body of findings on quality and to make recommendations.
- It will be important to include representatives of diverse communities as discussions continue.