



Issue Brief

May 2009

Publication #2009-13

OPRE Issue Brief #2



MULTIPLE PURPOSES FOR MEASURING QUALITY IN EARLY CHILDHOOD SETTINGS: IMPLICATIONS FOR COLLECTING AND COMMUNICATING INFORMATION ON QUALITY

Martha Zaslow, Kathryn Tout, Tamara Halle, and Nicole Forry

INTRODUCTION

As states and communities invest in initiatives to improve the quality¹ of early care and education, the measurement of quality is becoming more widespread and the importance of measuring quality well is gaining increasing attention (Zaslow, Tout, & Martinez-Beck, 2009). Within the broad context of interest in improving quality, this Issue Brief seeks to differentiate among a number of specific purposes for measuring quality in early childhood settings, and to identify the implications of these differing purposes for the careful and appropriate measurement of quality.

In this brief, we will:

- Review previous research that highlights the importance of identifying the purposes of measurement,
- Distinguish among different purposes for conducting assessments of quality in early childhood settings,
- Discuss the need for precaution when assessments seek to address multiple purposes at once, and
- Raise implications for developing future measures.

¹ For the purposes of this Issue Brief, we use the term “quality” to refer to the broad range of environmental features and interactions in nonparental care and education settings that have been positively linked to children’s development (Zaslow, Tout, & Martinez-Beck, 2009). These include structural characteristics such as child-adult ratio and the education of the teacher or caregiver, as well as process characteristics such as the frequency and tone of interactions between adults and children or activities that promote early literacy. The measures used to capture these dimensions of quality typically go beyond a focus on structural features to include a global assessment of process features as well as ratings of the daily routines and the physical environment. This conceptualization of quality differs from the standards used in child care licensing. While some states use well-known quality measures in their licensing systems as a way to establish a higher threshold of quality for programs, licensing typically establishes the presence of only a minimum level of basic health and safety routines and provisions. In this Issue Brief, our focus is on the measurement of quality above the floor established by typical licensing standards.

This Issue Brief series is supported under contract # HHSP233200500198U to Child Trends from the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, under the direction of project officer Ivelisse Martinez-Beck. Support is also provided by the Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. Martha Zaslow, Kathryn Tout, Tamara Halle, and Nicole Forry are researchers at Child Trends. The views represented in this brief are those of the authors and do not reflect the opinions of the Office of Planning, Research and Evaluation of the Administration for Children and Families.

PREVIOUS RESEARCH ON THE DIFFERING PURPOSES OF ASSESSMENT IN EARLY CHILDHOOD

To inform our discussion of the purposes of assessing quality in early childhood settings, it is illustrative to examine the related but distinct area of the assessment of developmental outcomes in young children. Distinguishing among different purposes of child assessments has provided a critical starting point for thinking through how assessments should be selected, how they should be used, and to whom information about assessment results should be communicated.

In *Principles and Recommendations for Early Childhood Assessments*, Shepard, Kagan, and Wurtz (1998) identified four intended purposes for assessments of young children: 1) to support learning, 2) to identify special needs, 3) to evaluate programs and monitor trends, and 4) to hold programs accountable. There are differences in the administration and use of child assessments based on these different purposes. For example, assessments to guide instruction are carried out in a familiar context where care and learning take place (the classroom or home-based care setting) by a familiar caregiver or educator. Assessments for this purpose are conducted on an ongoing basis and used to guide instruction for a particular child. In addition to being used by teachers, results are often communicated to parents.

In contrast, child assessments used for accountability help determine whether a school or district is meeting expectations, for example, by examining whether a targeted proportion of children in a school or district have reached a certain level identified as indicating proficiency. Consequently, measures used for this purpose need to meet high technical standards of reliability and validity, and they should be carried out in a standardized manner (for example, they are administered by a trained assessor at a particular time in the year). Individual child results are not reported; instead, results at the school or district level are communicated to policy makers and to the public. Shepard et al. (1998) cautioned against using a child assessment inappropriately for a purpose for which it was not designed or intended.

The more recent work of the National Research Council's Committee on Developmental Outcomes and Assessments for Young Children (Snow & Van Hemel, 2008) supports and extends Shepard et al.'s (1998) recommendations on the use of child

assessments. There are two key principles stressed throughout the NRC committee's report: That the selection of assessments and the way in which they are carried out need to be guided by the underlying purpose for which the measure was developed, and that early childhood assessments should not be carried out in isolation but should be part of a system with other key components. These further components include appropriate preparation of those who administer the assessments and those who interpret and use the information they produce, procedures to assure that child assessment results are interpreted in the context of knowledge about program quality and opportunities to learn, and advance planning for how needs for improvement will be addressed when they are identified. These authors also caution against using a measure to address multiple purposes. They propose that specific precautionary steps be taken when child assessments are carried out for multiple purposes.

DISTINGUISHING AMONG PURPOSES FOR ASSESSING QUALITY IN EARLY CHILDHOOD SETTINGS

Just as the identification of underlying purpose plays a central role in selecting, implementing, and communicating results from early childhood assessments, we propose that the assessment of quality in early childhood environments could be strengthened by articulating distinct purposes. Lambert (2003) comments on the need to differentiate among measures of quality. In his article, Lambert notes that measures of quality can be differentiated both in terms of intended recipient and breadth. Lambert notes that different recipients include programs themselves, researchers, or those determining whether a classroom or program has attained an externally determined standard of quality such as accreditation. These differing recipients may need information at different levels of detail. In terms of breadth, Lambert notes that some measures of quality focus on supports for specific domains of development, such as language and literacy development, while others provide a broad portrayal of overall quality. The purpose for measuring quality is critical to selecting specific measures. If the goal is for overall quality improvement, a broad measure may be most appropriate, whereas if the goal is to improve practice in a specific domain, a measure focusing in depth on a particular aspect of the environment may be more appropriate. In some instances, identifying the underlying purpose may call for the use of a combination of broad and domain-specific measures.

In this Issue Brief, we build on the articulation of the different purposes for early childhood assessment (Shepard et al., 1998; Snow & Van Hemel, 2008), as well as on Lambert's identification of differing goals for measuring quality in early childhood settings. In particular, we identify four different purposes for measuring quality in early childhood settings, and discuss the implications of these different purposes for the way in which data are collected, communicated, and used.

The four key purposes for measuring the quality of early childhood settings are:

1. To inform and guide improvement for individual practitioners or programs by identifying specific areas in need of strengthening.
2. To determine if program or policy investments have resulted in a change in quality over time, both at the level of the individual program and in a geographical area (such as community or state) where investments in quality have been made,
3. To contribute to knowledge about the contributors to and outcomes of quality, and
4. To describe or rate the quality of individual programs in a community or geographical area, with the aim of informing parental choice.

Just as the administration and communication of findings from child assessment measures differ according to the underlying purpose of their use, so too, measures of quality of the early childhood environment differ in terms of method of administration and communication of results according to these four purposes.

KEY DIFFERENCES BETWEEN THE PURPOSES FOR MEASURING QUALITY IN EARLY CHILDHOOD SETTINGS

The table at the end of this brief summarizes the similarities and differences in the use of quality measures for the four purposes outlined above according to several criteria:

- Who collects the information on quality,
- Who receives or uses the information on quality,
- How measures of quality are selected,

- The training requirements for those using the quality measure, and
- What supports are needed for the effective implementation of the measure for its intended purpose.

Below, we summarize the main distinctions among these four purposes according to each of these criteria.

COLLECTING INFORMATION ON QUALITY

The different purposes for quality measurement listed above require different skills and capabilities for collecting data. For example, technical skills for appropriate measurement are needed, but so are communication skills to inform and guide improvement by individual practitioners or programs (Purpose 1). When measurement of quality is conducted for this purpose, results need to be presented to providers in nonthreatening ways and used in creating plans to improve the practices of providers. When quality is assessed to evaluate whether change in quality has occurred in response to program or policy investments (Purpose 2) and to describe or rate a program's quality for the purpose of informing parents' choice of care (Purpose 4), it is critical to adhere to high standards of observer reliability. When data are collected to contribute to knowledge about the contributors to and outcomes of quality (Purpose 3), as well as when the goal is to rate program quality to inform parental choice (Purpose 4), data collection requires the capability to coordinate a large data collection effort and maintain high standards of reliability over time and across multiple data collectors, because data collection for these purposes involves multiple ratings over time and/or across geographical regions.

WHO RECEIVES THE INFORMATION AND HOW THE INFORMATION IS PRESENTED

How information on quality is received and how the information is presented depend on the purpose of assessment. For informing and guiding improvement by individual practitioners or programs (Purpose 1), care must be taken to present information from quality assessments in a constructive manner that can facilitate changes in the provider's practice (such as the nature and frequency of interaction with children), in the structuring of daily activities (such as how much time is spent in small vs. large group activities) and/or in the physical environment (such as the organization of space and the availability of materials for play and learning). Likewise, for communicating quality information to assist in parental choice of care (Purpose 4), it is important

that information be presented in a way that is easy for families to comprehend and that provides the information parents find most useful in making child care choices (for example, the most useful information may include both an overall summary rating as well as information on key components so that parents who find one particular aspect of quality most important have separate information on this aspect). For assessing the effectiveness of quality investments or understanding the antecedents and outcomes of quality (Purposes 2 and 3), results are provided in more technical reports to funding agencies, policy makers, and researchers. Though funding agencies and policy makers are the primary audience for results of assessments conducted for Purpose 2, and researchers are the primary audience for Purpose 3, all three groups may be included in dissemination efforts of quality measurement carried out with these purposes in order to help coordinate research, funding decisions, and policy approaches.

SELECTING MEASURES

Across the four purposes of the use of quality measures, an overarching theme is the need to select the measure of quality in keeping with the aspect or aspects of quality of greatest interest. Those conducting quality assessments may select administrative document reviews, surveys, or observational instruments. Among observational instruments, there is variation regarding what is being measured: an overall or global measure of quality, a measure of fidelity to a particular curriculum, or an in-depth focus on a specific aspect of quality. Two factors have led to an increased focus on measures of specific aspects of quality such as stimulation in the early childhood setting for language and literacy development or the development of early math skills: recent research documenting modest associations between global measures of child care quality and child outcomes and a heightened sensitivity to the potential of early education/care settings as a foundation for later academic achievement.

TRAINING

Reliability in conducting quality assessments is an overarching theme across the four purposes of assessment. Continuous and consistent adherence to stringent standards for reliability is particularly important for evaluating program and policy assessments (Purpose 2), assessing the associations between factors contributing to quality and the aspects of quality associated with positive child outcomes (Purpose 3), and rating programs in a participating geographic area to inform parental choice (Purpose 4). Not only initial training but also

ongoing oversight may be necessary to ensure that reliability is both established and maintained. For Purpose 1, additional training is needed for guiding individual program improvement in order to prepare assessors to present results of quality assessments to providers and to use these results in quality improvement plans.

Some states and localities that use quality measures in their systems offer a range of training on the quality measures to different stakeholders. For example, training may be abbreviated to facilitate familiarity or it may be more in-depth to increase understanding of the measures by those who will need to use the information to guide improvement efforts. Trainings targeted to different stakeholder groups may contribute to a higher comfort level with the measures and greater buy-in to the measures among stakeholders, and protect against interpretations that are not supported by the measures.

IMPLEMENTATION

Quality measurement, particularly when it is conducted on a broad scale, may require an infrastructure to oversee implementation activities. Implementation involves making key decisions about all aspects of data collection and dissemination, including:

- Which measures are used,
- Which programs are observed and how often,
- How document review is carried out and verified,
- Whether administrative data can be collected on an ongoing basis,
- How to support and supervise those who observe quality and provide technical assistance on quality improvement,
- How to assure that the measurement of quality maintains standards of reliability system-wide,
- How to address questions and concerns from programs, and
- How to assure ongoing dissemination of the quality information to its intended audiences.

The priorities for establishing an infrastructure for quality measurement implementation will differ somewhat by purpose. For guiding improvement by individual providers and programs (Purpose 1), the hiring and supervision of staff who specialize in developing quality improvement plans and providing feedback on progress based on quality

assessments are key issues. The development of an infrastructure for ensuring reliability in measurement on an ongoing basis is a key issue for Purposes 2, 3, and 4. Verifying information gathered from providers and programs is relevant to determining if program investments have resulted in a change in quality (Purpose 2), as well as to providing quality ratings to parents (Purpose 4). Finally, finding the best ways to present quality information in nontechnical ways is a key issue for Purpose 4 when parents are the target audience, and for Purpose 2, when funders and policy makers are the target audience.

PRECAUTIONS REGARDING USING MEASURES OF QUALITY FOR MULTIPLE PURPOSES

Initiatives at the state and local level experience pressure to use one data collection effort to measure quality for multiple purposes. This is especially the case given the expense of reliably collecting data on quality in many early childhood settings. Compared to multiple data collection efforts, a single data collection effort providing data for multiple purposes has the benefits of being more efficient and avoiding the potential of overburdening early care and education settings. Yet just as in the discussions on using a single early childhood assessment for multiple purposes, it is important to anticipate that collecting data on quality for multiple purposes in a single data collection effort runs the risk of failing to attend to important considerations regarding measures selection, reliability, communication, or infrastructure needs for a particular purpose. We are beginning to see thoughtful consideration of these issues in state and local data collection efforts.

Below we note four issues that are arising as communities and states collect data on quality for multiple purposes. For each issue, we also note practices that are being put in place in selected states or communities to address the issue. It should be noted that the field is at an early stage of identifying specific precautions for using one data collection protocol to address multiple purposes of quality measurement. Thus, the issues noted below and the examples of precautionary practices to address the issues should be considered a starting point to be built upon and extended over time.

ISSUE # 1

The use of differing standards for reliability when data are collected to inform consumers, evaluate

the outcomes of quality investments, and guide quality improvement efforts by individual providers or programs

Most Quality Rating Systems (QRSs) implement both a rating process and a quality improvement process for the providers who participate in the system. When this is the case, information on quality may be used to inform consumers (Purpose 4). However, it may also be used as a source of information to guide improvement in individual programs (Purpose 1) and summarized to inform policy makers regarding whether investments in quality are resulting in overall progress (Purpose 2). Although reliability standards historically have been less stringent when quality information is collected to guide improvements in individual programs, states are now taking precautions to use stringent standards for reliability when a single round of data collection will be used to inform individual programs, consumers, and policy makers.

For example, some states allow only data collectors who have demonstrated initial and sustained adherence to strict reliability standards to collect data that contribute to the quality rating in a QRS (Purpose 4) and the technical assistance process used to help providers improve the quality of their programs (Purpose 1). The trained data collector then provides the results of the quality measurement to contribute to an overall quality rating, but also shares the information with the technical assistance specialist who has the training necessary to help guide goal-setting and improvement strategies based on the results of the measure.

This practice eliminates the possibility that the data collector and the technical assistance specialist would score the measure differently. It also honors the separate and critical expertise of technical assistance providers who have specialized skills for helping providers through the quality improvement process (Thornburg et al., in press). Finally, this practice respects the need to have stringent standards for reliability when measurement is used for accountability in evaluating the outcomes of quality improvement investments across programs. Information that is trusted by the public and decision makers is a high priority for Purposes 2 and 4.

ISSUE # 2

The use of information on quality for research and generalizable knowledge when the initial intent was to collect ratings to inform parents

As states and localities with measurement systems accumulate data across programs and over time,

this information may be sought as a data source for generalizable knowledge, that is research that seeks to inform the general understanding of the contributors to and outcomes of quality (Purpose 3). For example, these data could be used to identify predictors of quality or the characteristics of providers who make improvements over time compared to those who do not. One consideration states are confronting in this area is that when data are collected for generalizable knowledge, specific planning steps need to be taken before rather than after the data are collected. These planning steps include confirming that a specific measure of quality has been validated for the purpose of research and following human research subjects protection procedures (for example, obtaining approval from an Institutional Review Board and informed consent from participants).

Lambert (2003) notes that different measures of quality may have been validated only for the purpose of improving program quality or only for research purposes. In some cases, different versions of measures have been developed for various purposes of measurement. Even if precautions are in place to protect the privacy of data on quality, (for example, using identification numbers rather than names and reporting only aggregate results), these protections may not suffice if the information is going to be published in a journal article or research brief as a way to contribute to generalizable knowledge. When data are disseminated for generalizable knowledge, participants in the data collection need to be aware in advance of the intended use and choose to participate. States and localities may want to consider collecting informed consent with all quality data if they anticipate using the results for these broader purposes. They may also want to review measures of quality to confirm that there is evidence of validity for the multiple purposes being considered.

ISSUE # 3

The use of information on quality to evaluate public investments in quality without adequate information on change over time or contextual factors, and without resources to implement a plan for quality improvement

In its volume on early childhood assessment, the Committee on Developmental Outcomes and Assessment of Young Children (Snow & Van Hemel, 2008) recommends that data collected for accountability purposes be complemented by data collected on corresponding contextual factors, such as the demographic risks for children and families and the resources available to follow up on children

identified as at risk for developmental problems. Such information can be critical in explaining findings and informing targets for investments in improving child outcomes.

This recommendation can be extended to data on the quality of early childhood programs. For example, understanding that lower-quality ratings tend to occur in programs with limited access to professional development opportunities can provide information to shape future investments and prevent high-stakes decisions (such as de-funding) from being implemented without full understanding of the context.

The Committee on Developmental Outcomes and Assessment of Young Children (Snow & Van Hemel, 2008) also calls for using data on children's progress over time rather than just point-in-time assessments. In a similar manner, it may be more helpful in a measurement system focusing on quality to identify where improvements are and are not occurring, and to seek to understand what is contributing to different patterns of change over time. Supplementing quality data collected at multiple time points with information on program resources and investments in quality, along with demographic characteristics of the families served and other contextual information, will make it possible to identify factors contributing to changes in quality.

ISSUE # 4

Presentation of information on quality to multiple audiences without adequate background information and explanations of measures

A final problem states and localities are encountering is one in which data collected for one purpose and one audience are shared more broadly with other stakeholders. For example, reports designed for a research audience may be disseminated to parents, providers, and policy makers who may lack background information on the measures being used or who do not have the background to understand technical research language.

The Committee on Developmental Outcomes and Assessment of Young Children (Snow & Van Hemel, 2008) cautioned that part of implementing a system of child assessments involves providing key stakeholders with information about the measures and how to interpret scores so findings are understood appropriately and misinterpretations are avoided. In a parallel manner, states are finding that it is important to plan for and provide information to all key stakeholder groups in the appropriate interpretation of quality measures.

Communicating about quality data is especially important if the information may influence decisions by particular stakeholders.

These four issues are examples of the kinds of situations states and communities are beginning to encounter when using information on quality for multiple purposes. It is encouraging to see the emergence of precautionary steps for each of these scenarios.

SUMMARY AND IMPLICATIONS

The measurement of quality in early care and education settings is expanding as states and communities launch initiatives to strengthen quality. While there may be a common, underlying concern with strengthening quality, there are nevertheless important distinctions in the more specific purposes for the collection of data about quality. This Issue Brief has identified four different purposes for measuring quality in early care and education settings: to guide improvement by individual providers and programs, to determine if program and policy investments have resulted in improvements in quality at the level of individual programs or multiple programs in a geographical area, to build knowledge about what factors contribute to quality and what aspects of the environment contribute to specific child outcomes, and to describe or rate the quality of individual programs to inform parental choice.

Following the precedent of work on the assessment of development in young children, we note that the purpose underlying assessment of quality in early childhood settings has important implications for

what data are collected, how data are collected, and how results are communicated. In this brief, we have highlighted some key similarities and differences in the selection, training, administration, and dissemination of quality data across these four purposes. These differences in data collection and use underscore the importance of planning a data collection effort with clarity about the underlying purpose.

We have also highlighted the pressure that states and communities are under to collect quality data in an efficient and cost-effective manner, which often leads to a single data collection being used for multiple purposes. We have outlined four issues that states and communities may face and suggested precautionary action to guard against the misuse of quality measures when a data collection effort addresses multiple purposes. Those involved in this field should continue discussing the additional problems states and localities are confronting and the precautions that are needed when measurement of quality is carried out for multiple purposes.

In addition, as further issues are identified in which measures of quality that are initially collected for one purpose are serving a second or third purpose, it will be useful for the field to turn to measures developers to ask them to clarify the intended uses of measures. Measures developers could provide specific guidance on both the appropriate use of the measure, with precautions in place as needed, and the purposes for which the measures should *not* be used, even with precautions put in place. This guidance will be useful as increasing weight is put on certain measures to serve multiple roles in a measurement system.

REFERENCES

- Lambert, R. G. (2003). Considering purpose and intended use when making evaluations of assessments: A response to Dickinson. *Educational Researcher*, 32(4), 23-26.
- Shepard, L., Kagan, S., & Wurtz, E. (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: National Education Goals Panel.
- Snow, C., & Van Hemel, S. (2008). *Early childhood assessment: Why, what and how?* Report of the Committee on Developmental Outcomes and Assessments for Young Children. Washington DC: National Academies Press.
- Thornburg, K. R., Mauzy, D., Mayfield, W., Scott, J.L., Sparks, A., Mumford, J., et al. (in press). Data-driven decision making in preparation for large-scale QRS implementation. In Zaslow, M., Tout, K., Halle, T., & Martinez-Beck, I. (Eds.), *Next steps in the measurement of quality in early childhood settings*. Baltimore: Brookes Publishing.
- Zaslow, M., Tout, K., & Martinez-Beck, I. (Manuscript submitted for review). Measuring the quality of early care and education programs at the intersection of research, policy and practice.

The information contained in this brief was drawn from a paper produced as an outgrowth of participation in the Developing the Next Wave of Quality Measures for Early Childhood and School-Age Programs meeting that was held January 23-25, 2008, in Washington, DC. A summary of this meeting will be available on the Child Care and Early Education Research Connections website, <http://www.childcareresearch.org>.

*A more detailed versions of this paper will be part of a forthcoming book entitled **Next Steps in the Measurement of Quality in Early Childhood Settings** which will be published by Brookes Publishing, Baltimore, MD.*

Purpose	Who Collects the Information on Quality?	Who Receives/Uses the Information on Quality?	Selecting Measures	Training	Implementation	Emerging Issues
<p>Purpose #1: To inform and guide improvement by individual practitioners or programs by identifying specific areas that need to be strengthened.</p>	<p>Information on quality usually is collected by an individual providing technical assistance to the early childhood caregiver/ educator or program.</p> <p>Alternatively, a provider may be trained to collect quality information on his/ her own program, or an outside observer may complete the quality assessment and then provide information to the person providing technical assistance.</p>	<p>Information on quality is provided to individual provider/early educator or lead and assistant teacher jointly for a class or group. The program director may also receive this information.</p> <p>The intent of providing this information is to develop a plan for quality improvement and document whether improvement has occurred/been sustained.</p>	<p>The measure needs to align with aspects of quality that the provider or program is seeking to improve.</p> <p>For example, the goal may be to improve overall quality, in which case a global measure is most appropriate. Alternatively, fidelity measures would be used for assessing the implementation of a curriculum and more detailed measures would be used when quality in one domain (e.g., health) is being assessed.</p>	<p>Training needs to focus not only on reliable collection of quality data, but on translating specific quality indicators into guidance for program improvement.</p> <p>Training also needs to focus on forming a relationship with the provider and supporting program improvement.¹</p>	<p>The availability of those experienced in early childhood education who are also skilled at supporting other providers/ teachers in making changes in their programs is an issue.</p> <p>Implementation of technical assistance approaches requires not just initial preparation of those providing technical assistance, but thinking through a process for ongoing supervision and support for these professionals. Thus, having adequate numbers of qualified staff to provide on-site technical assistance as well as adequate supervision are key issues.</p> <p>Tools are needed for tracking implementation of quality efforts with sites.</p>	<p>A key emerging issue here is whether the reliability standards set for evaluation/or research purposes need to be applied to the assessments of quality for this purpose.</p> <p>There is some evidence that those providing technical assistance do not rate providers as stringently as those collecting data for evaluation and research purposes.</p> <p>In some states with quality improvement efforts, this issue is being addressed by increasing the rigor of training on quality measures for those providing technical assistance to providers. As an alternative, some states have had one data collector obtain data for both technical assistance and research purposes.</p>

¹ There is increasing discussion and sharing of appropriate approaches for training of those providing technical assistance. For example, NACCRRA has developed best practice standards on issues like caseload and background. The Partnerships for Inclusion consultation model evaluated through the QUINCE Evaluation and Early Childhood Educator Professional Development Programs are beginning a process for sharing manuals for professional development of coaches. Note lack of agreement on terminology—technical assistance, coaching, mentoring, facilitation.

Purpose	Who Collects the Information on Quality?	Who Receives/Uses the Information on Quality?	Selecting Measures	Training	Implementation	Emerging Issues
<p>Purpose #2: To determine if program or policy investments have resulted in a change in quality over time, both at the level of the individual program and in a geographical area (such as a community or state) where investments in quality have been made.</p>	<p>To assure independence of the evaluation, observers need to be independent from the programs observed.</p> <p>Observers for specific evaluation studies are often part of a research team. When possible in experimental studies, it is desirable for observers to be unaware of whether a program has received an intervention.</p>	<p>Information is provided to funding agency and/or policy makers.</p> <p>Information on quality is used to guide decisions about whether a program or initiative is continued, expanded, or modified.</p> <p>Information may also be disseminated to researchers and the public via technical reports or journal articles.</p>	<p>Selection of quality measure(s) should align closely with the goals of the initiative or program.</p> <p>For example, if the initiative has a goal of broad improvement in overall quality, a broad observational measure might be selected, whereas if the goal is to improve language and literacy practices or instructional quality, a different measure might be selected.</p> <p>Further development of measures addressing certain aspects of quality is underway.</p> <p>The use of multiple measures for assessing different aspects of quality may be appropriate, as well as interviews with staff and program document reviews.</p>	<p>Training is needed that will permit observers to obtain and then maintain stringent requirements for reliability in completing observational measures.</p> <p>Assessments of raters' reliability in using the tool should be conducted both before they start rating programs and periodically throughout the course of data collection.</p> <p>Ensuring accurate, reliable measurements are critical because results have potential consequences for maintaining, expanding, or discontinuing programs/initiatives.</p>	<p>An infrastructure is needed for ongoing observations and/or document review and interviews. This is particularly important in statewide evaluations with multiple sites (e.g., Quality Rating Systems).</p> <p>Verification of data provided through interview and document review help ensure the accuracy of collected data. Verification of data can be done using registries, existing sources of verified information, or requests for further documentation from programs.</p>	<p>Best practices for designing infrastructures for the collection of reliable data and the verification of reported data are needed.</p>

Purpose	Who Collects the Information on Quality?	Who Receives/Uses the Information on Quality?	Selecting Measures	Training	Implementation	Emerging Issues
<p>Purpose #3: To contribute to knowledge about the contributors to and outcomes of quality.</p>	<p>Data on quality with this goal is usually collected within the context of a longitudinal research study.</p> <p>Observations may be collected by multiple collaborating university research teams or by a survey research firm with the capability to conduct direct observations of quality across multiple sites.</p>	<p>Information on quality is shared with researchers with the goal of building the knowledge base to strengthen quality or child outcomes.</p> <p>To strengthen policies and programs, descriptive research results also need to be communicated to the practice and policy communities.²</p>	<p>Alignment is needed between the selected measure and the aspect of quality that is being assessed/goal of the study.</p> <p>For example, if the goal of a study is to consider how early childhood environments foster development in specific domains, the measure of quality needs to provide sufficient detail to explore possible contributors to development in this domain.</p> <p>Further measures development work is in process to address the need for greater specificity in documenting some aspects of the environment.</p>	<p>Observers should be trained to obtain and maintain reliability in accordance with the requirements set forth by measure developers.</p>	<p>When a study is carried out at multiple sites, procedures need to be developed to assure the consistency of observation collection and coding practices across sites.</p>	<p>Recent findings point to statistically significant but modest relationships between widely used measures of quality and child outcomes. These are challenging researchers to develop measures focusing in greater detail on specific aspects of the environment that, in turn, are hypothesized to be related to children's development in specific domains.</p> <p>Gaps in current measurement tools are found in particular age groups (e.g., infants and toddlers) and settings (e.g., home-based care). Additionally, measures are needed that focus in depth on particular aspects of quality.</p>

² A recent example of communication of descriptive research results to research but also policy and practice communities includes the research reexamining the role of the bachelor's degree as a predictor of quality in early childhood settings. The Research Connections (www.researchconnections.org) and NCCIC (www.nccic.org) websites are potential sites for sharing information with the practice and policy communities.

Purpose	Who Collects the Information on Quality?	Who Receives/Uses the Information on Quality?	Selecting Measures	Training	Implementation	Emerging Issues
<p>Purpose #4: To describe or rate the quality of individual programs in a community or geographical area, with the aim of informing parental choice.</p>	<p>This purpose requires a group of data collectors and infrastructure for collection of quality data on an ongoing basis in a designated geographical area.</p> <p>Some states have partnered with universities or community organizations to collect the data. Other states have created a unit within their licensing or human services departments.</p>	<p>Summary ratings of quality can be provided to parents to inform choices; these ratings can also be provided to policy makers so they can identify types of early care and education settings and aspects of quality that need to be improved.</p>	<p>States vary in their data collection for this purpose. Some states use surveys and document reviews to record structural aspects of quality, while others use observational tools. Additionally, some states have tiered systems in which observations are reserved for providers who have met higher standards of quality.</p> <p>States using observations have generally employed measures of global quality, though there is an emerging trend toward adding or substituting with measures of quality that have a more explicit focus on aspects of quality that support early learning.</p> <p>A key issue when multiple measures are integrated into a summary rating is how to weight the different components.</p>	<p>Training needs to provide a basis for data collectors to obtain and then maintain stringent standards for reliability.</p> <p>Issues of reliability pertain to recording of administrative data (document reviews), too.</p>	<p>There are four primary implementation issues for this purpose of assessment. First, because assessments in this purpose can have consequences for providers in terms of public perception and possible enrollment if parents use quality rating systems to choose their care, high standards of reliability must be maintained for both observational measures and document reviews.</p> <p>Second, explicit appeal processes should be set up and documentation of measurement practices and standards should be clear.</p> <p>Third, the frequency of ratings must be determined. The ratings should be current so parents can rely on them as they make child care decisions.</p> <p>Finally, to reach an overall rating of quality for the program, decisions must be made regarding the number of classrooms to observe and the process for selecting those classrooms.</p>	<p>Ratings for this purpose will occur across various types of child care. To date, few measurement instruments have been developed for family, friend, and neighbor care or for care provided to infants and toddlers. The reliability of data derived from observational measures and/or document reviews is also central.</p> <p>Finally, information is needed on how to provide information in an accessible and easy to use format. Unanswered questions include: whether parents find summary ratings or component ratings (focused on particular facets of care) more useful, whether quality ratings are being communicated effectively to parents of differing cultural and economic groups, and whether there are constraints that limit the capacity of parents to use quality information.</p>