Washington State Early Learning and Development Benchmarks

Evaluation Plan

Authors:
Pia Rebello Britto, Ph.D.
Sharon Lynn Kagan, Ed.D.
Kristie Kauerz, M.A
Kate Tarrant

National Center for Children and Families, Teachers College, Columbia University

February 2005

Attachment 5.2.4

Written Reports of efforts to assess
Voluntary Early Learning Guidelines

Underlying the process of developing the Washington State Early Learning and Development Benchmarks (Benchmarks) was a fervent commitment to inclusivity, objectivity, and creativity. Underlying the content of the Benchmarks was a commitment to their comprehensiveness; scientific basis; age, cultural and linguistic appropriateness; and relevance to current contexts in which children in Washington State live. To ascertain the degree to which these commitments have been realized, an evaluation is necessary.

In general, evaluations are a systematic way of assessing an initiative or project to ensure that intended goals and objectives are met. Three different kinds of evaluations may be used to meet these purposes for the Benchmarks. First, a process evaluation can assess the degree to which the process of developing the Benchmarks met its intended goals. Second, a validity evaluation can ensure the accuracy of the Benchmarks' content and age-expectations for young children in Washington. Third, periodic evaluations of the content and use of the Benchmarks can assure that the effort continues to meet its goals (or changed goals) over time. In sum, three distinct types of evaluation plans for the Washington State Early Learning and Development Benchmarks are presented below:

I. Process Evaluation
II. Validity Evaluation
III. Periodic Evaluation

I. PROCESS EVALUATION

Rationale

Before the Benchmarks were developed, key stakeholders in Washington created a comprehensive list of principles to guide the development process. The primary purpose of a process evaluation then is to examine adherence of the actual process to the intended principles. Such an evaluation can help to assess the efficiency, effectiveness, and accountability of the process used to develop the Benchmarks. A process evaluation can also provide stakeholders with an understanding of what worked well and what could have been improved upon during the development process.

Goal

The goal of a process evaluation is to: [a] ascertain the fidelity of the actual process of developing the Washington's Benchmarks to the intended process; [b] discern which, if any, steps in the process were essential to the ultimate outcome of the document produced; and [c] investigate the efficiency and effectiveness of the development process. Based on the stated principles, a set of criteria are recommended for process evaluation:

Table: Recommended Criteria for the Process Evaluation

Evaluation Criteria	*Examples of Evaluation Questions*
Inclusivity	How inclusive was the process of development? Were key stakeholders involved at all stages? Was expert, caregiver, and parental opinion considered?
Expansion and Alignment with Existing Resources	Did the process build on existing documents and research in the field?
Efficiency	How long did it take to complete the development process? Were deadlines met in a timely manner? Did the timeline allow for a sufficient, but not excessive, number of meetings? Did the timeline allow for a sufficient review?
Collegiality and Collaboration	How did all participants work together? Was the process of development respectful and inclusive of all points of view? Did members of the development team attend and participate in meetings? Were conflicts resolved in an amicable manner? Whenever possible, was consensus achieved?
Capacity Building	Did the members of the development team gain knowledge and skills during the process of development? In addition to contributing their individual expertise, did team members learn from the process?
Future Efforts	What were the key lessons learned from the process that could be applied to other efforts in Washington State?

Methodology

Two potential approaches can be used to conduct a process evaluation: (1) Survey Research or (2) Process Validation:

Survey Research is a means of systematically collecting data or information by asking a set of pre-formulated questions in a predetermined sequence in a questionnaire or interview to a set of individuals. Survey questions cover a range of topics, such as asking people about their knowledge, attitudes, preferences, behaviors, and/or opinions. Survey research implies the collection of new data.

Process Validation is an approach used to compare the actual process of development to the stated or intended process. While often used in scientific fields such as pharmaceutical research, in which the critical elements and procedures in the manufacturing process of pharmaceutical products are evaluated to ensure that the process operated effectively and within the intended protocol of development, this approach can be easily adapted for evaluating the Benchmarks' development process.

Option 1: Survey Research

Surveys are commonly used to gather systematic information from a sample of individuals on a particular topic of interest. A survey can be conducted in many ways (e.g., on-line, phone, in-person, mail) and structured in many ways (e.g., pre-selected response options, open ended questions). Defining the format and structure of a survey is important, as these characteristics influence the nature and pertinence of the data collected. In addition to considering the format and structure, a survey must be standardized so that the same set of questions is given to each respondent and the questions are asked in the same manner. Standardization ensures that the aggregate responses describe the group's perspective rather than a compilation of unrelated individual opinions.

The advantage of using survey research for the process evaluation is that it is inclusive and data can be obtained from a diverse set of constituents. A disadvantage, however, is that this method is time and resource dependent; designing, conducting, and analyzing survey data takes time and requires a high level of methodological expertise.

If this option is selected, given the goals of the Benchmarks work, we recommend: (a) on-line; (b) confidential; (c) structured survey that should be sent to a (d) representative sample of individuals who participated in the development process. Each of these components is defined below.

(a) *On-line*: A survey that can be completed using the technological advantages of the internet should be developed. Benefits of an on-line survey include cost-effectiveness and confidentiality of responses. In situations where the respondents do not have access to a computer, a hard copy version of the survey should be sent to the respondents via mail and a stamped, addressed envelope included.

(b) *Confidential*: No identifying information (e.g., name, address, phone number) should be solicited from the respondents. Information regarding a professional profile (e.g., job category, urban/rural/suburban location) could be requested so as to understand the results better. Confidentiality is important as it allows respondents the freedom to answer the questions without fear of any consequences from their responses.

(c) *Structured*: The questions in the survey may be open-ended (e.g., allow for free-form responses) or close-ended (e.g., require a choice from a pre-determined range of responses). The advantage of open-ended questions is that they add great texture to the responses, capturing details that might be lost with closed or structured questions. The disadvantage of open-ended questions is that they require a detailed analysis, with coding procedures clearly delineated. Close-ended questions come in several types (multiple-choice, categorical, ordinal, numerical, and Likert-Scale).

The Benchmarks evaluation may utilize binary-categorical response (yes/no) questions if the responses do not need to be nuanced (see example below). If the questions require a wider range of responses, the response options could be expanded to solicit agree-disagree responses along a continuum, called a Likert-Scale question (see example below). The advantage of close-ended, as opposed to open-ended, questions is that the responses may be easily tallied and, therefore, results may be obtained relatively quickly. It is quite typical to use a combination of open-ended and close-ended questions when the information sought cannot be obtained accurately with one format only.

Example: Binary-Categorical Questions

	Yes ***1***	***No*** ***0***
Did you feel your opinion mattered?		
Were the working groups conducted in an efficient manner?		
Did you gain any new knowledge and skills during the development process?		

Example: Likert-Scale Questions

	Strongly Agree ***5***	***Mildly Agree*** ***4***	***Neither Agree nor disagree*** ***3***	***Mildly Disagree*** ***2***	***Strongly Disagree*** ***1***
The process of development included representation from all relevant stakeholders and users					
The working groups were conducted in an efficient manner					
I gained new knowledge and skills during the development process					

(d) *Representative Sample*: A representative sample should be selected t complete the survey. The sample should be selected from participants in the development process (e.g., Core Team, Advisory Panel, Affinity Groups, Review Groups, and families, parents, and caregivers). Given that the aim of the process evaluation is to determine fidelity of the actual process to the intended process, a few members from each of the participating groups should be randomly selected to respond to the survey.

Option 2: Process Validation

Less resource dependent and more efficient than survey research, process validation is effective when an understanding of the development procedures is more important than inclusivity of respondents in the analysis. Like survey research, process validation compares what ***was*** done to what was ***intended*** to be done. Process Validation can be conducted in two different ways. The first is the prospective method, where the system to test the process is set up prior to the actual development. The second is the retrospective method which looks back at the process, once it has been completed. Given that the Benchmarks are in the final stages of development, the prospective method is not appropriate and, therefore, an outline for conducting the retrospective approach is presented.

(a) *Identification of Review Committee*: Three or more individuals should be invited to form the review committee. In order to ensure an unbiased evaluation, the review committee should include representatives from three groups:

Group 1: People who helped craft the guiding principles *and* were involved in the development process itself;
Group 2: People who were *not* involved in crafting the guiding principles and development strategy but were involved in the actual development process; and
Group 3: An independent scientific methodologist who did not participate in crafting the guiding principles and did not participate in the actual development process.

(b) *Drafting of the Validation Protocol:* The review committee should draft a validation protocol identifying the central components of development process. The list of central components is derived from the development strategy and guiding principles. In a validation protocol, the central components are defined in terms of their identifying characteristics and illustrative examples (see example below). Fidelity to the process is determined based on rating these central components or key constructs. Obtaining accurate data for the evaluation requires concise definitions that are clearly understood by the entire review committee.

Example: Validation Protocol

Central Components	*Definition*	*Examples*
Built on prior work	The process utilized existing information and evidence in developing the Benchmarks	1.Head Start Child Outcomes Framework 2. Preventing Reading Difficulties in Young Children (Snow, Burns & Griffin, 1998); and so on…
Inclusivity	Key stakeholders were involved in the process of development	1. Early Child care providers 2. Parents 3. Early Childhood Experts; and so on….

(c) *Data Collection*: The review committee should then collect all public-use documentation related to the development of the Benchmarks, including published reports, guidelines, inter-office memos, meeting agendas, drafts of the documents, and group e-mail correspondence.

(d) *Data Coding and Analyses:* Based on the validation protocol, the data are coded vis-à-vis the degree to which the actual process varied from the intended process is analyzed. The primary goal of coding the data is to rate the alignment between the intended process and the actual process (see example below). Coding of the central components requires that each observable aspect of the central component is listed in a measurable form. Then the data are entered for each of these aspects. Finally, the data are coded. Prior to coding the data, the review committee must establish a set of criteria for coding, so as to ensure internal reliability (e.g., a score of "2" means the same thing to all members on the committee). To illustrate using the example below, the review committee would determine the codes for an aspect of a central component, such as 0 = no documents reviewed; 1 = a few documents reviewed; 2 = several documents reviewed, but the documents were not relevant; 3 = several relevant documents reviewed, though not a comprehensive coverage; 4 = most relevant documents reviewed; 5 = all relevant documents reviewed.

Example: Coding a Central Component

Central Component	*Definition*	*Examples*	*Measurable and observable aspect*	*Data Coding (0 = no documents reviewed; 5 = all documents reviewed)*
Built on prior work	The process utilized existing information and evidence in developing the Benchmarks	1.Head Start Child Outcomes Framework 2. Preventing Reading Difficulties in Young Children (Snow, Burns & Griffin, 1998) and so on…	1. Number of scientific texts reviewed:____ 2.Number of States' Early Learning Standards reviewed:____ and so on…	1 2 3 4 5

(e) *Generation of Results:* All the scores generated from coding the data are tallied to determine fidelity of the development process to the intended development strategy. In addition, the results should specify processes that were particularly useful and make recommendations to improve the quality and efficacy of other similar initiatives.

Guidelines and Special Considerations

Listed below are several considerations to address and guidelines to be kept in mind when designing the process evaluation. These considerations and guidelines are equally applicable to both approaches – survey research and process validation.

Reliable and valid evaluation data: Conducting a sound and robust process evaluation involves the collection of data that are relevant and required. It is essential that study questions be crafted, or central components defined, to ensure that the data collected are comprehensive and accurately address the goals of the evaluation.

Evaluation should be useful: It is important that the results from the process evaluation are generated in a timely manner and are informative for the stakeholders. The evaluation report should be disseminated to, and easily understood by, the intended audience.

Protection of participant rights: Given the nature of a process evaluation, a "no-blame" strategy needs to be agreed upon in writing. The evaluators need to respect the worth and dignity of those who were involved in the various stages of the process of development and any conflict of interest should be treated openly and honestly.

Balanced reporting: The strengths and weaknesses of the process of development should be evaluated in an unbiased manner and reported comprehensively and fairly. All factors that

might have affected the process of development need to be acknowledged when reporting the results, whether they were directly evaluated or not.

Next Steps

An effective and comprehensive plan to conduct a process evaluation of the Washington State Early Learning and Development Benchmarks could be launched by:

- *Deciding on the evaluation design.* Based on the key evaluation questions, resource availability, and timelines, one of the two methods presented above should be selected. For instance, if resources are limited and there is a tight timeline for producing the results, then the process validation method might be more appealing. If fidelity to the inclusivity of the process is most important, then survey research may be desirable. Criteria that should be considered when selecting a method include the feasibility, practicality, and cost-effectiveness of each option. Ultimately, the chosen evaluation procedure should be minimally disruptive to the participants and the benefits of the evaluation should justify the resource expenditure.

- *Identifying a leadership group to oversee the process evaluation.* Someone needs to be vested with the responsibility for coordinating the process evaluation. This leadership team should be familiar with data collection methodologies so as to be able to select an appropriate evaluator to design the evaluation, collect data, and analyze the obtained data in a manner that ensures reliable and valid results.

II. VALIDITY EVALUATION

Rationale

In response to one of the primary aims of the Benchmarks ("to promote reasonable expectations and practical strategies for parents and others who care for and teach young children to support their learning and development"), the original Benchmarks document provides indicators of development and learning, by age group, for each of the goal statements (see Benchmarks document for details). These indicators were developed based on careful reviews of developmental literature, institutional and agency guidelines (e.g., American Academy of Pediatrics), and standards from other states within the United States and nations around the world. Although the indicators were based on sound theoretical foundations and a comprehensive review of the literature, they were not tested for content and age appropriateness for children in Washington State. While there is no reason to expect that Washington State varies systematically from other states and regions in the U.S., scientific prudence suggests that the age and content of the Benchmarks should be validated to be certain they are appropriate for children in the State of Washington.

Goal

The goal of the age and content validity evaluation is to determine if the content of the indicators and the age-level expectations are appropriate for children residing in the State of Washington. That is, the goal is to discern if the content of the Benchmarks, as they are currently written, are applicable and appropriate and if the age expectations are too easy, too hard, or just right for the indicated ages of the children.

Methodology

Content and age appropriate validity of the Benchmarks refers to the extent to which the Benchmarks accurately reflect expectations for what children in Washington should know and be able to do within the specified age ranges. Presented in this section are options for conducting age appropriate and content validity evaluations.

Age Appropriate Validity Evaluation: The purpose of the age appropriateness evaluation is to determine whether the age-level expectations of the Benchmarks are accurate and valid for children from birth to entry-into-Kindergarten residing in Washington State. A *descriptive quantitative study* is necessary to establish the age validity of the Benchmarks in a scientifically acceptable manner. While this methodology will provide numerical evidence for the age alignment of the indicators, it is heavily dependent upon both staff and financial resources.

Content Validity Evaluation: The purpose of a content validity evaluation is to determine if the content of the Benchmarks accurately represents the expectations for young children's learning and development. It represents an additional check to be sure that the content of the proposed Benchmarks is appropriate for children in Washington State. A *focus group qualitative study* is proposed for the content validation. A focus group is an organized discussion with a selected group of individuals to gain information about their views, opinions, and experiences on a particular topic or issue. The strength of this methodology is that it is designed to understand people and the social and cultural contexts within which they live from the point of view of the participants. Such information can be lost when textual data are quantified in a quantitative methodology. A disadvantage, however, of the qualitative focus group methodology is that the results obtained from the group discussions may require an additional level of expert analyses in order to realign the indicators to reflect the desired content.

1. Age Appropriate Validity Evaluation

A quantitative age validity study might include the following steps:

- *Determine the age groups to be used.* In the Benchmarks document, four age groupings are listed (Group 1 = birth to 18 months; Group 2 = 18 to 36 months; Group 3 = 36 to 60 months; and Group 4 = 60 months to entry to Kindergarten).

Children should be selected from the upper end of each of the age groupings for the evaluation, as that is the age by which the specific behavior is expected to be mastered.

- *Determine the percentages of achievement.* Decisions need to be made regarding cut-off percentages at which the indicator would be considered normatively age-appropriate (e.g., 40%? 60%?). If 90% of the children can achieve the indicator, it is no longer an expectation to be achieved; then age expectation (anticipated age of accomplishment for normatively developing children) would need to be lowered. If, on the other hand, 20% of the group achieved the indicator, it might be considered too difficult and so the age expectation would need to be increased.

- *Devise a sampling plan that yields a representative sample of children.* The sampling frame should be developed based on: (1) representation from all possible groups of children in Washington State to whom the Benchmarks might be applied; (2) the sampling requirements of the statistical analyses to be conducted; (3) resource availability; and (4) timeline.

- *Devise the data collection plan and finalize the instrument for the evaluation.* The evaluation instrument should be a direct observation of children's behavior, using a sample of the Benchmark items. The items selected for the observation instrument should be amenable to direct observation and represent the range of behaviors to be evaluated.

- *Data Collection and Analyses.* Data should be collected using the instrument in scientifically reliable ways with trained observers. The data should be checked for accuracy. Primary data analyses techniques involve conducting statistical analyses such as obtaining percentages, frequencies, means and correlations. The results from these analyses provide information on the age appropriateness of the indicators.

- *Re-Aligning Indicators.* Based on the results obtained, the indicators may need to be re-aligned to reflect accurate age and developmental expectations.

2. *Content Validity Evaluation*

A list of criteria to analyze the content of the Benchmarks is presented below. These criteria are based on prior work focused on content analyses of early learning standards (Kagan & Britto, 2004).

Table: Recommended Criteria for Content Validity Evaluation

Evaluation Criteria	*Examples of Evaluation Questions*
Breadth	Are all domains of early child development represented in the Benchmarks?
Balance	Are the domains represented in comparatively equal distribution?
Depth	Are there a sufficient number of indicators in each domain to reflect its scope comprehensively?
Accuracy	Do the indicators accurately reflect the domain?
Hierarchical	Are the indicators listed in a developmental order, i.e, subsequent indicators build upon the previous indicators?
Cultural inclusion	Has sufficient attention been paid to cultural diversity?
Alignment with national and state standards	To what extent do the Benchmarks link with other State and National Standards?

The following steps are suggested for conducting the focus group sessions:

- *Select the focus group participants.*
  - First, a recommended group size for a single focus group is approximately 10 people so as to ensure that all participants have their voices heard and the group is manageable.
  - Second, participants to a group should be selected based on homogeneity of experience and area of expertise. Even though it would seem that diversity in the group might be beneficial, heterogeneity in groups does not appear to be very effective. For instance, if one of the participants is very senior in an agency and the other participants are not, the less senior participants may be hesitant to express their individual opinion in front of the senior member. The group composition needs to be balanced to ensure that all participants will feel comfortable expressing their opinions. Therefore, it is recommended that the group be more homogenous (i.e., participants with similar characteristics be invited).
  - Third, participants should represent early childhood educators, parents, and early childhood experts.
  - Fourth, the relationships among focus group participants should be transparent. Putting people together who know each other should be avoided as personal relationships might bias the results.

- *Select a moderator.* It is important that the moderator be able to facilitate the sessions effectively so as to elicit the opinions of all the participants and collect content validity information relevant to the evaluation. The moderator should be able to follow a set of guidelines to maintain consistency across groups; the guidelines should not make the moderator rigid in his/her approach.

- *Distribution of the benchmarks.* After the moderator has been appointed and the focus group participants have agreed to participate, they all need to receive a copy of the Benchmarks at least two weeks prior to the focus group session. A the success of a focus group session is largely dependent upon the level of familiarity the participants have with the document prior.

- *Procedure for the Focus Group Session*
    - *Structure.* The structure of the focus group has a strong influence on the data obtained so that there needs to be some flexibility in the group to enable the free expression of opinions. The conversation should be guided to flow in a direction that reflects the group's values, without getting off-track or derailed.
    - *Questions.* The questions should be asked in a conversational manner, in-sequence, clearly stated, open-ended, and neutral. Questions should address both issues related to the overall Benchmarks document (e.g., Are the goal statements, indicators, and strategies culturally relevant? Are the Benchmarks an accurate reflection of what you expect that children in this age group should know and be able to do?) and issues related to each specific domain (e.g., Does the information for each domain reflect the local values and expectations? What could be add/revised/ to make the domain more relevant?).

- *Data collection and analyses.* Various methods of data collection are used to capture the information gained in focus group session -- video tape, audio tape, and manual note taking. Each method presents unique advantages and disadvantages. For instance, videotaping the session facilitates comprehensive data collection that can be viewed and coded at a later point in time. However videotaping is expensive and the coding is labor intensive. Manual note taking, on the other hand, may not be a very reliable source of data collection, as it is dependent on the skills of the note-taker. However an advantage of note-taking is that is inexpensive and relatively easy to code. In addition to considering the data collection technique, data analyses should focus on the patterns, themes, and perspectives that emerge from the focus groups.

Guidelines and Special Considerations

The following general guidelines are suggested to aid in developing the age appropriate and content validity evaluations.

- *Age Appropriate Validity Evaluation.* First, the evaluation should be representative of the children residing in Washington State. It is important that this evaluation reflect the appropriateness of the indicators for the youngest citizens of the state. The cultural, racial/ethnic, linguistic, ability, regional and socio-economic diversity in the State of Washington needs to be taken into consideration when devising a sampling frame that is representative of all children residing in the state.

Second, the age groupings selected for the study should map on exactly to the age groupings of the Benchmarks. If the age-ranges between the evaluation and the benchmarks are not aligned, the results of the evaluation will not be valid for the Benchmarks

- *Content Validity Evaluation.* The participants in the focus groups should represent multiple perspectives, including parents, experts, and educators.

Next Steps

An effective and comprehensive plan to evaluate the age appropriateness and the content validity of the Washington State Early Learning and Development Benchmarks could be launched by:

- *Deciding on the evaluation design, format and questions.* Based on the key evaluation questions, resource availability, and timelines, the evaluation needs to be designed. Criteria to be considered while crafting the design include feasibility, practicality, and cost-effectiveness. Ultimately, the evaluation procedure should be minimally disruptive to the participants and the benefits of the evaluation should justify the resource expenditure

- *Identifying a leadership group to oversee validity evaluations.* Someone needs to be vested with the responsibility for coordinating the validity evaluation. This leadership team should be familiar with data collection methodologies so as to be able to select an appropriate evaluator to design the evaluation, collect data, and analyze the obtained data in a manner that ensures reliable and valid results.

III. PERIODIC EVALUATION

Rationale

"Change" is a dominant force in society today. Whether it is greater racial/ethnic or cultural diversity, the educational policy landscape, or technological advancement, change is rapidly occurring across all facets of life. In order for any initiative or program to be successful, it must be current and keep up with changes that affect its implementation and effectiveness. Outdated programs lose their utility as they become either no longer appropriate for the times or the fit between the program and the population it serves is mismatched. The Benchmarks were developed taking into consideration the social, economic, and education climate of 2004-05. However, several years from now, the content of the Benchmarks may not be as relevant for young children in Washington State. For this reason, the Benchmarks should undergo periodic evaluations to ensure that the content is relevant and useful for the current times. Furthermore, one of the guiding principles for the Benchmarks states "that the benchmarks will be reviewed

and updated every 5 years." Consequently, it is recommended that the Benchmarks should be reviewed and updated every five years.

Goals

The goal of the periodic evaluation is to ensure that the content of the Benchmarks remains current and that the Benchmarks are being used effectively to support the learning and development of young children in Washington State.

Methodology

A periodic evaluation is a widely used method to review and update programs, policies, and products. The advantage of a periodic evaluation is that it facilitates longevity of established and successful initiatives, ensures that the operating mechanisms and outputs are up-to-date, and provides the opportunity to reassess and re-evaluate effectiveness. A periodic review of the Benchmarks' content will ensure that the statements of expectation, for what children should know and be able to do, are current, relevant, and reflect the most recent research on young children's development. Another important purpose of the periodic evaluation is to ensure that the Benchmarks are being used for their stated purposes (e.g., they have not outlived their utility, and that they are still considered to be functional framework for improving the lives of young children).

Listed below are a set of criteria that need to be considered in conducting the periodic evaluations.

Table: Criteria for Content Periodic Evaluation

Criteria	*Examples of Evaluation Questions*
Relevant for demographic landscape of society	Are the Benchmarks still relevant for the children residing in Washington State? Do the Benchmarks capture the state's socio-economic, racial-ethnic, and linguistic diversity?
In keeping with the latest scientific knowledge	Are the Benchmarks based on the latest literature and knowledge base? Are they based on the most recent developments in early childhood practice?
In keeping with changes in lifestyles	Do the Benchmarks reflect the environments in which young children spend long periods of time? Do the Benchmarks reflect the needs of families caring for young children?
Concordance with Educational Policies	Are the Benchmarks aligned with early childhood policies?
Alignment to other educational standards	Are the Benchmarks still closely aligned to other educational standards in the state? Are the Benchmarks closely aligned with other national early childhood standards?

Table: Criteria for Use Periodic Evaluation

Criteria	*Examples of Evaluation Questions*
Current Users	Who are the current users of the Benchmarks? Are they primarily parents? Early childhood educators? Are they parenting programs?
Purpose	What is the primary purpose for which the Benchmarks are being used? Are the Benchmarks being used in parenting programs? Are they being used for teacher preparation programs?
Most Useful Aspects	Have the goal statements been useful in developing uses for the Benchmarks? Have users found learning strategies helpful?

A general procedural is outlined for conducting the content and use periodic evaluations of the Benchmarks.

- *Formation of a Review Committee.* Four and half years after the benchmarks have been implemented, a review committee should be formed to carry out the Benchmark review. The review committee should consist of 15 to 20 members who are experts in early childhood and early learning standards and who represent key stakeholders, some of whom are not employed by the sponsoring agencies.

- *Formation of a Leadership Core.* The review committee should nominate a group of 5 to 6 individuals to form the leadership core of the committee. The primary function of the leadership core is to provide the guidance for the review and ensure that the review is completed in a timely manner.

- *Determine Focus of the Review.* The leadership core needs to articulate the focus of the review. The specific purposes for reviewing the content of the Benchmarks and evaluating their use need to be spelt out clearly. The specific purposes then need to be ratified by the entire review committee.

- *Review and Update the Guiding Principles.* The leadership core should review the existing guiding principles to ascertain their relevance to the periodic review. Principles that do not appear to be relevant need to be either adapted or substituted with appropriate principles. Additional principles may be added (e.g., the intention of the review should be constructive; the review process should be conducted in a congenial manner; and the recommendations should be based on documented evidence).

- *Method for the Review Process.* The method for the review process needs to be guided by a set of procedural parameters:

  (i) Content Review. A sub-committee should be formed from the overall review committee to oversee the content review of the Benchmarks. The focus group methodology described above could be employed for the review.

(ii) Use Review. Another sub-committee should be formed from the overall review committee to oversee the use review of the Benchmarks. The sub-committee should review data collected over the past 5 years on how the Benchmarks have been used.

- *Results and recommendations.* The sub-committees then need to analyze the results of their respective reviews and generate a set of recommendations for updating the Benchmarks. The analyses must include a discussion of the strengths, weaknesses, and a rationale for the recommendations. The entire review committee should ratify the recommendations.

- *Adopting the Recommended Updates*. The review committee then decides by consensus the recommendations that will be adopted in the updated version of the Benchmarks. Finally, the review committee also needs to approve the recommendations taking into consideration the current educational policies, demographics of Washington State residents, and potential new uses for the Benchmarks.

Guidelines and Special Considerations

The following guidelines are suggested to aid in the development of the content and use periodic review evaluations:

- *Importance of Periodic Review Established:* It is important to acknowledge the existence of the periodic review and its role in ensuring that the sustainability of the Benchmarks. By doing so, everyone involved with the Benchmarks initiative knows from the outset that this review will take place and the Benchmarks evolve as they are updated.

- *Fair, Unbiased, Independent and Inclusive Review*: The review committee needs to include members who are independent of the sponsoring agencies and members who were not involved in the development of the Benchmarks, such as key stakeholders, users, and key constituents involved in the implementation of the benchmarks. The review process needs to be inclusive of the opinions of all members of the review committee.

- *Timetable for review:* It is important that timelines for deliverables and Benchmark updates are communicated to stakeholders, and users of the Benchmarks. The timelines and expectations for the update need to be realistic and take into consideration the steps and time involved in the review process.

- *Generation of Data:* The data utilized in a periodic evaluation is typically collected during the cycle of implementation. Therefore, as the Benchmarks are being implemented, data collection needs to be strategically planned to ensure that

information needed by the review committee to conduct the periodic evaluation will be available.

Next Steps

The following steps will need to be completed in order for the periodic evaluations to be conducted in a timely manner:

- *Allocation of Resources*: As the plans are being drawn up for the implementation of the Benchmarks, and concomitant expenditures, resources need to be allocated for the periodic evaluation. It is important to budget for this evaluation in advance so as to ensure it will take place. Additionally, if there is no data collection component in the implementation plans, then resources need to be allocated to the review committee in order for them to undertake an evaluation of the implementation to inform the revision of the Benchmarks.

- *Assurances of Continuity*: There need to be assurances in place that future implementation of the Benchmarks will depend upon revised and updated benchmarks, thereby ensuring that the review will be conducted.