

**State Justice Communities Planning Initiative**

**Evaluation Instrument**

**Report on Pilot Tests**

**June 19, 2003**

**Greacen Associates, LLC  
HCR 78 Box 23  
Regina, New Mexico 87046  
505-289-2164  
[john@greacen.net](mailto:john@greacen.net)**

## Summary

This report summarizes two of the three pilot tests of the seventh draft of the State Justice Communities Planning Initiative Evaluation Instrument. It includes some observations concerning the third test. It records the deliberations of the Design Team in evaluating the pilot test experience and recommending final changes to the evaluation instrument and the process by which it is administered.

The two completed pilot tests took place in Washington and Kentucky. A third test has been conducted in Ohio. Full results, including consensus scores, from that test are not yet available for inclusion within this report.

This report describes the pilot tests in Washington and Kentucky, including the composition of the evaluation team, the process used in preparing for and conducting the week long site visit in each state, how the instrument was scored and how the scores were reported to the participating states. It also includes reports on the process used in the Ohio test.

Appendix A to the report sets forth the individual and group consensus scores for Washington and Kentucky. The wide disparity in the initial scores assigned by the different team members shows that the instrument is not suitable for use by an individual evaluator. The consensus scoring process, however, appears to be valid and methodologically supportable, given that both the Washington and Kentucky teams were able to agree on a consensus score for every item, even when that score differed from the initial scores of every member of the team. We recommend that the instrument be used with groups of at least three evaluators, using the consensus scoring process developed during the pilot tests in Washington and Kentucky.

Appendix B compares the scores for Washington and Kentucky. The results show that the instrument differentiates among different state processes and accurately reflects the state planning process in each state.

The report analyzes a number of issues concerning the evaluation process and the instrument, making a series of recommendations for revision and enhancement of both. Some key issues – such as the size of the evaluation team, the length of the site visit, and travel by team members during the visit – were tested in the Ohio pilot, and observations concerning that experience are included in the discussion.

## Table of Contents

Summary .....	1
Table of Contents .....	2
Table of Contents .....	2
Background .....	3
The Pilot Tests .....	3
Inter-rater Reliability-Team Member and Team Consensus Scoring .....	7
Subjective Validity of the Evaluation Results .....	11
Issues With Respect to the Evaluation Process .....	11
Preparation .....	12
Initial Scoring of the Instrument Based Only on the State Plan .....	12
Travel in the Course of the Onsite Evaluation .....	13
Length of the Evaluation Site Visit .....	14
Composition of the Evaluation Team .....	15
Maintaining a Record of the Evaluation .....	15
The Access Database .....	16
The Format and Content of the Final Report .....	16
Use of the Interview Protocol .....	17
Issues With Respect to the Instrument .....	17
Recognizing the Collective Nature of the Evaluation Process .....	17
Giving Credit for Activities That Are Occurring But Are Not Included in the Plan....	17
Allowing the Use of a Broader Range of Scores .....	18
Changing the Labeling of the Parts of the Instrument .....	18
Clarifying the Language in the Instrument .....	19
Separately Scoring the Plan and Its Implementation .....	19
Revising the Method for Scoring Section (now Part) 2-Implementation .....	20
Section (now Part) 3-The Objective Measures .....	20
A1 through A3-Statewide Capacities.....	20
A4-Non LSC Resources Received by All Legal Services Providers Serving Persons Eligible for LSC-funded Services .....	21
A5-Relative Availability of Legal Services Lawyers .....	21
B1-Quantity of Service Provided by Type of Service .....	22
C1-Geographic Equity in Resource Distribution .....	23
C2-Equity in the Provision of Services to Groups of Clients .....	23
D1-Annual Costs of State Planning .....	25
Additional Questions Identified in the Preparation Guide.....	25

## **Background**

Greacen Associates, LLC has been assisting the Legal Services Corporation and its Design Team to devise an instrument for use in evaluating state level planning for the provision of legal services to the poor – the LSC State Justice Communities Planning Initiative.

The instrument has gone through seven drafts. In the fall of 2002 it was circulated for comment to legal services programs throughout the country. Revisions were made in the instrument to reflect some of the comments submitted. Action on many of the comments was deferred pending field testing of the instrument.

The instrument has now been tested in three pilot states – Washington, Kentucky and Ohio. This report summarizes the experience in the first two pilot tests and preliminary observations concerning the final pilot test in Ohio and makes recommendations for changes to the instrument and to the process by which it is administered. The recommendations reflect the deliberations of the Design Team, which met to discuss a preliminary draft of this report on May 28 and 29, 2003.

John Greacen from Greacen Associates participated in the first two pilot tests. It is those tests for which full information is available as of the date of this report.

This summary includes comments from the LSC state planning team. It takes into account the responses from Washington and Kentucky to the evaluation reports prepared by the evaluation teams administering the instrument in their states. No such report has yet been prepared for Ohio.

## **The Pilot Tests**

The instrument was tested in Washington during the week of January 6<sup>th</sup>, in Kentucky during the week of January 27<sup>th</sup>, and in Ohio during the week of May 12<sup>th</sup>.

All three states were informed that the pilot tests were conducted for the purpose of testing the instrument, and did not constitute formal evaluations of the states' justice communities planning efforts. However, the evaluation was otherwise conducted as if it were a formal evaluation, the instrument was scored, and a report was prepared and sent to the pilot state for review and comment. Each state also received a detailed questionnaire eliciting its views and experiences with the evaluation process.

All three pilot tests employed the seventh draft of the instrument. The evaluation team for Washington used a printed version of the instrument. The Kentucky pilot used an access data base for recording the scores of the team members. The Ohio test used a reformatted, printed version of the seventh draft of the instrument.

For all three pilot tests, the state was provided the instrument well in advance of the site visit. Each state also received a “preparation guide” detailing LSC’s expectations of the state for the evaluation process. Both LSC and the state designated “principal contacts” to work together to plan for each site visit.

The principal contacts developed the schedule for the site visit and the list of persons to be interviewed. The LSC principal contact developed an initial list of implementation actions (derived from the state plan) on which the evaluation team would focus; the state had an opportunity to add action items to that list. (Kentucky availed itself of that opportunity.) The state principal contact assembled data for the evaluation team bearing on the topics in the instrument.

Written materials – including various versions of the state plan -- were distributed in electronic and written form to the team members prior to the site visit. These materials were supplemented during the site visit by the state principal contact in response to inquiries from the evaluation team and by persons whom the team members interviewed.

Both Washington and Kentucky extended every courtesy and the fullest possible cooperation to the evaluation teams. The processes and interactions in both states were invariably cordial, positive and constructive.

The evaluation team assembled on the evening before the evaluation began, met with the state principal contact and other support persons with whom they would work during the week, and went over the agenda for the week. The team then met by themselves to make interview assignments, to discuss their own procedures, and to do more detailed planning for how they would conduct the scoring of the instrument and prepare an exit report for the state.

Each visit began and ended with meetings with a cadre of state legal services leaders. On the first morning, the evaluation team explained the purpose of the pilot tests (to test the instrument and not to conduct an official evaluation of the state planning effort), explained the process it would follow, and invited the state leaders to provide background on the state’s legal services structure, planning process, and legal services history. (During the Kentucky pilot, the leaders were asked specifically to state the greatest benefit that the state had obtained from the state planning effort and the major challenges remaining.) The exit meeting provided a summary of the team’s findings and recommendations (but no specific scores) – clearly noting that these findings and recommendations represented the views of the individual team members and not the views of the Legal Services Corporation. The state leadership group was invited to respond to the findings and recommendations and to make observations about the evaluation process.

Each team designated one of the Design Team members as the team leader for purposes of guiding its discussions and serving as its principal spokesperson during the entrance and exit meetings with the state legal services leadership.

The Washington and Kentucky teams spent three and a half days conducting a large number of interviews and visits to programs. The Ohio team had closer to two days for interviews. In Washington and Kentucky, the team members traveled throughout the state conducting interviews; in Ohio they remained in one location. In all three sites team members conducted some interviews by telephone. In every state, at least one interview did not take place as planned because the interviewee did not appear or the team ran out of time for interviews. The team members made a sincere effort to meet before and after each day's interviews to share the information and observations that each of them had developed during the day. That process was intended to give each team member the same information upon which to score the instrument at week's end.

In Washington and Kentucky, the evaluation team asked persons interviewed how much time they had devoted to preparation for and conduct of the interview. The team also asked the state to provide an estimate of the cost of the whole evaluation process. Persons interviewed usually responded that they had spent very little time preparing for the interview. None of the pilot teams made an effort to add up the number of days expended for interviews and neither state was able to provide an estimate of its overall evaluation costs.

In Washington and Kentucky, the team members scored the instrument individually on the evening of the second to last day of the visit. They met the next morning to reconcile their scores – developing a single consensus score for the team to report. For every scored item, the team members discussed the reasons behind their individual scores and, invariably, were able to reach a consensus team score. There was no instance during the Washington or Kentucky scoring that the team members reached an impasse on a consensus score. The individual and team consensus scores were recorded (although the identity of the team member contributing each individual score was not recorded).

In Ohio, the individual team members did not score the instrument before beginning their joint deliberations. The scoring process did not go as well in Ohio. The team did not complete Section 1 of the instrument before they turned their attention to the preparation of themes to report to the state leadership cadre during the exit interview.

The evaluation reports were to be prepared by the LSC principal contact. The process of producing those reports was protracted. John Greacen ultimately prepared a draft report for Kentucky to serve as an example of how the instrument could be used to serve as the vehicle for conveying the results. Reports for Washington and Kentucky were then prepared by the LSC principal contacts following that model. The team members were given a short period of time to review and comment on the draft reports. Few of the team members commented, perhaps because of the extended time that had elapsed – over two months from the end of the visits.

Washington and Kentucky both responded to their reports.

The pilot tests differed in some regards. The Washington site visit team consisted of four members – two members of the Design Team, an LSC state planning team member, and the consultant. The LSC state planning team member for Washington was not able to participate in the Washington site visit. Another LSC state planning team member substituted for her. He did not have the in-depth history of Washington's state planning efforts that the person for whom he substituted would have brought to the process. Randi Youells joined the team for the Friday scoring process as an observer. She and one of the other team members met with the Washington Access to Justice Board the following Monday and provided an abbreviated summary of the team's exit report.

The exit reports did not include any specific evaluation scores. Instead they consisted of general observations and recommendations from the individual team members (clearly stated not to be the positions of LSC). These observations and recommendations became the basis for the narrative summaries and recommendations sections of the final evaluation reports for both states.

The Washington site visit took six days, beginning on Sunday afternoon and ending midafternoon on Friday. The Washington test involved over 80 interviews. Most of those interviews were conducted by a single team member, though, when possible within the interview schedule, two team members participated in interviews. The interviewers took notes on the interviews, using a standard interview protocol prepared in advance of the site visit. As the week progressed, interviewers relied less and less on the interview protocols – focusing the interviews on the key questions that were emerging for the team as a whole. The team did not interview any of the three senior state legal services planning staff, although they met often with them to discuss the process of the visit and their needs for additional information.

In Washington, the team members traveled moderately. Two team members flew from Seattle to Spokane, spent a day conducting interviews there, drove to Wenatchee, spent the night in Wenatchee, met with legal services and community members in Wenatchee, and drove back to Seattle. Other team members drove to Olympia for interviews.

Although the team had intended to score the instrument individually Thursday afternoon and begin team scoring on Thursday evening, the late arrival of the team members from Wenatchee delayed that schedule. They met on Thursday evening to share information, scored the instrument individually Thursday night, and did their consensus scoring and prepared their exit report on Friday morning. They presented the report after lunch on Friday afternoon and did not reassemble for further scoring. One of the team members had not scored the individual actions for the implementation section of the instrument, so the team did not address Section 2 of the instrument. It also did not address the objective measures of Section 3 except for parts A1 through A3.

In Kentucky, the site visit team consisted of five members – two Design Team members, the LSC state planning team member for Kentucky, the newly-hired LSC staff person responsible for state planning evaluation, and the consultant. The team followed

the same six day schedule used in Washington. The team conducted roughly 70 interviews, often with two team members participating in the same interview. The team used the interview protocol in the same fashion as it had been used in Washington – relying upon it less and less as the week progressed. The team interviewed one of the two senior legal services planning staff members; Kentucky objected to the team's failure to interview the other. One of the team members interviewed her by phone the following week.

All team members traveled extensively throughout Kentucky. Two members traveled to Western Kentucky for a day and a half. Two traveled to Louisville for a day. Two traveled to Appalachia and North Central Kentucky for two days. Individuals also made other day trips from Lexington, the site visit headquarters, to visit the state capital in Frankfort and other program sites for interviews. The team made a strong effort to share all of the information it developed, but was handicapped in doing so by the extensive travel. It too was unable to begin team scoring before Friday morning.

The Kentucky team attempted to use the Access data base. Four of the five members recorded their scores using the data base. The fifth used a printed copy of the instrument. The consultant was unable to generate a report from the Access data base summarizing the scores of the four who used it. So, the team proceeded with the scoring using a printed copy of the instrument, with each team member reading off his or her scores from the Access data base. The team did not have time to complete the scoring of Section 2 or Section 3. It held three subsequent conference calls of roughly two hours each to complete the scoring of Section 2. It did not complete Section 3 beyond Parts A1 through A3.

The Ohio test was designed to test the use of the instrument by an exclusively LSC staff team, with a smaller team and a shorter schedule. The Ohio test included three team members, a three day schedule and no travel outside Columbus for interviews. The team conducted about 40 interviews, many of them by telephone. The interviews focused on specific aspects of the plan and its implementation. Consequently, one of the senior legal services leaders was interviewed three times. The Ohio team did not score the instrument individually before beginning their consensus scoring. The team did not complete the scoring of Section 1 of the instrument and did no scoring of Sections 2 or 3 while on site. One team member was assigned responsibility for suggesting scores for the remainder of the instrument.

## **Inter-rater Reliability-Team Member and Team Consensus Scoring**

Appendix A displays all of the individual scores reported by the team members in Washington and Kentucky, as well as the team consensus scores for Section 1 and Section 3A1-A3. For Kentucky, it also displays the scores for Section 2 – the more than 80 implementation items rated by the Kentucky team. The way the scores were recorded, it is not possible to attribute a particular score to a particular team member.

Consequently, it is not possible to determine whether particular team members recorded scores consistently above or below those of other team members. In Washington, the team members did not observe any such consistent pattern. In Kentucky, two of the members commented that they seemed to be consistently lower than other team members.

Question marks appear whenever a team member felt s/he did not have enough information to make a scoring judgment. In two instances in the Washington pilot, the entire team concluded that it did not have sufficient information to assign a score.

In some instances, a midpoint score is recorded (e.g., 3.5). This is used when a team member reported inability to decide between two scores (in this instance, 3 or 4). The team as a whole chose a score of 3.5 for the overall score for Section 2 of the Kentucky instrument, not being able to agree on 3 or 4.

None of the Kentucky team members had independently decided on an overall score for Section 2. Interestingly, the result was that the team took longer and had a more difficult time reaching consensus on this score than on any action item score. This incident demonstrated the importance of the individual scoring process for the success of the team consensus scoring.

It is immediately obvious that there is extreme variability within the individual scoring. For some items, scores ranged from 5 to 1 – the maximum possible variation among the individual scores. More striking, individuals' scores for some items were evenly distributed across the entire range. (For subparts 3A1 to 3A3 the maximum variability was 3 to 0; in some instances, individual scores crossed that entire range, too.) It was rare indeed that all team members agreed on a score. In rare instances, the consensus score was different from any of the individual scores. For instance, all individual scores may have been high, but after discussion the team decided the score should actually be very low.

The Washington scores appear somewhat more consistent than those for Kentucky. The Section 2 (implementation) scores for Kentucky seem to be somewhat more consistent than the scores for Sections 1 and 3, although there are instances in which those scores are very inconsistent as well.

The overall scores tended to be more consistent than the component scores. The aggregate scores for Sections and Parts tended to be more consistent than the overall scores for Subparts. In sum, the more global the issue being scored, the more consistent the scores became.

Greacen Associates attributes the wide variation in individual scores to four factors – differences in the information available to the individual team members, differences in the judgmental processes of the different team members, exhaustion and information overload, and the persistent inability of the team members to maintain the distinction between the plan and the reality of legal services delivery in the state.

Although both the Washington and Kentucky teams attempted to share the information they developed, it was not possible to repeat every comment or observation. Those persons who conducted an interview, or viewed a demonstration, had a stronger impression of that interaction than others whose experience was vicarious. Consequently, some of the difference in individual scores arose from differences in the information available to the scorer, and from differences in the weight given to a particular observation, statement or opinion.

The questions posed in the instrument often call for subjective judgments. For instance, item 1.1.1 asks whether a state's process is "highly client centered," "adequately client centered," or "not client centered." Answering that question not only calls for marshalling a large amount of information gathered from the plan and from the site visit, but it also calls for the exercise of the rater's judgment, based on the rater's own experience and expectations. It was clear from the consensus scoring discussions that each team member brought a different perspective to these judgment questions. However, when those differences in perspective surfaced, all team members were invariably able to subordinate their personal judgments to the collective analysis of the team as a group.

The individual scoring was done at night on the fifth day of a long, intense week. The consultant spent four hours on the scoring process in each state. Other team members reported needing a comparable length of time. This meant that team members were working until midnight on their individual scores. The process involved marshalling large amounts of information from diverse sources. Particularly in Washington, the team was provided mountains of paper to absorb. It was not clear to the consultant that all team members were able to digest all of the information provided, or that they paid attention to the most important items. In particular, at least one team member did not seem aware of the contents of the state plan itself during the scoring at one of the sites.

Finally, both the Washington and Kentucky teams encountered serious difficulty applying the instrument as intended when the implementation of the plan did not match the high objectives set forth in the plan. The design of the instrument calls for the assessments in Section 1 to be based on the plan itself, supplemented with information on activities performed by the state that may not appear explicitly in the plan. A state should get full credit for considering an issue fully in its plan, whether or not it has been able to implement that section of its plan effectively or fully. Both the Washington and Kentucky team members had great difficulty maintaining this differentiation. As a result, they tended to give the state lower scores in Section 1 than the plan itself deserved. They "marked a state down" for weaknesses in its structure or service delivery, even though the state plan identified that area for improvement and contained a solid strategy for addressing it.

During the scoring of the Washington evaluation, the entire team was affected by this phenomenon. In reflecting on the process, several members realized what they had done. The consultant reported the phenomenon to LSC and to the members of the

Kentucky team. All agreed that the instrument should be scored as intended. The Kentucky team discussed the issue in depth twice – during the initial meeting on Sunday afternoon and on Thursday evening immediately before they adjourned to score the instrument individually. Yet the team encountered the same phenomenon repeatedly during the consensus scoring process. The consultant, having experienced it in Washington, was able to draw attention to the issue when it arose and the team members were able to apply the instrument appropriately when they realized that their scoring of the planning process was being affected by their negative opinions of plan implementation.

For example, the team was shocked to learn that all four of Kentucky's legal services programs – between them -- had only one black lawyer. Kentucky was aware of its poor minority hiring record and had convened a task force to study diversity in its work force. The task force produced a good report, calling for enhanced training, recruitment and retention of minority staff. All four programs agreed to implement the task force report and have been doing so. But this is a recent development and there is still only one black legal services lawyer in the state. The Kentucky team members found it almost impossible to give Kentucky full credit for planning for diversity, even though the state had done a commendable job of investigating the problem and designing a sophisticated course of action to correct it. After considerable discussion, the team was able to agree on a fair score. But the issue arose again when the team members commented on the completed report for Kentucky. This issue – the persistent tendency to reduce the planning scores because of implementation failures -- remains a serious one for fair, effective and consistent application of the instrument. It is addressed again later in this report.

It is clear from these results that the instrument is not suitable for use by a single evaluator. However, because of the success of the teams in reaching consensus on team scores as a result of the collaborative consensus scoring process, the instrument appears to be effective when scored by a group. Consequently, Greacen Associates recommends that the instrument not be used by teams smaller than three persons. We recommend that the team members score the instrument individually and then use a consensus process to discuss the rationale for their individual scores, deriving consensus scores through that process. The consensus scores should not be derived by averaging the individual scores, but rather by discussing them thoroughly, so that the team remains open to a consensus score different from any of the individual scores.

The pilot testing process does not tell us whether two separate groups of evaluators, evaluating the same state at the same time based on the same materials, interviews and demonstrations, and using the consensus scoring process, would produce the same set of scores. Such a test is not feasible. LSC does not have the resources for such a duplicative test; nor could two different teams have the same experience in the same state at roughly the same time. Greacen Associates is of the opinion that the collective scoring process is sound, and that the consensus scoring process will produce consistent and fair results. Our conclusion is bolstered by consultation with a social science researcher in New Mexico who advised us that the professional consensus

scoring process is an appropriate method for combining quantitative and qualitative perspectives to form valid evaluative conclusions.

## **Subjective Validity of the Evaluation Results**

Appendix B displays the consensus scores for Washington and Kentucky for the purpose of allowing LSC staff and the members of the Design Team to make a judgment whether the instrument is valid in the sense that its results accurately represent the sophistication of state justice communities planning in the two states and that those results differentiate between the two states – reflecting relative strengths and weaknesses of the two state’s processes, plans, and implementation efforts.

Greacen Associates’ assessment is that the instrument has shown that it is capable of capturing accurate and complete information about the first two test states, and that it differentiates between them appropriately. That view is based on participation in the first two site visits and on the favorable reports from Washington and Kentucky on the results of the pilot tests. Although each state noted a score or two that it considered inaccurate, both states reported that the report reflected the reality of state justice communities planning and the strengths and weaknesses of the legal services delivery systems in their states. Further, they found the summary comments and recommendations of the evaluation team insightful and helpful.

We do note that the scores for Washington on Section 1 of the instrument were probably lower than they should have been in some instances, given the failure of the evaluation team to consistently differentiate between the plan and its implementation.

That the summary scores for the two states on Section 1 are the same reflects the reality of what the teams observed – that Washington’s planning effort has significant weaknesses that are not generally recognized, and that Kentucky’s planning effort is surprisingly robust and sophisticated – well beyond what the national legal services community perceives.

The Design Team concurred in the consultant’s judgment that the instrument produced accurate scores, based on the reports from Washington and Kentucky and from the consultant’s personal observations. One member noted that the scores for these two states tended more towards “4”s and “5”s than “2”s and “1”s, but that this result was predictable and appropriate given the choice of two states with sophisticated planning processes for the initial pilot tests.

## **Issues With Respect to the Evaluation Process**

The Design Team revisited a number of issues concerning the state justice planning initiatives evaluation process during their May 28-29 meeting in Chicago.

## ***Preparation***

Both state's questionnaire responses suggest that a state needs more time to prepare for the evaluation than they had. The Design Team recommends that LSC adopt a practice of giving at least three months advance notice to a state of an evaluation.

The preparation manual needs to be expanded to cover:

- agreement between the state and LSC on the timing of the evaluation visit. The Kentucky visit came during a critically important legislative session, imposing particular pressures on state legal services leaders and creating sensitivities about interviews that would not have existed if the legislature were not in session.
- a pre-evaluation orientation session for key state legal services personnel. It would be helpful for the LSC principal contact and the LSC staff member in charge of state planning evaluation to meet with the state principal contact and other key leaders well in advance of the site visit to go over the process and the instrument in detail.
- agreement on persons to be interviewed. In the Ohio test, the evaluation team identified a series of topics on which they wished to focus, asking the state to identify specific persons knowledgeable about each topic. More attention needs to be placed on the identification of the right persons to interview and on how LSC and the state reach agreement on the list. In no instance, however, should the same person be interviewed more than once.
- suggesting to the state that it arrange for demonstrations of particularly important programs or automated systems that have resulted from statewide planning.
- LSC should take more initiative in identifying and providing documents critical to the evaluation. For instance, LSC has copies of all versions of a state's plan and its self-assessment. LSC should not have to rely on the state to provide these materials to the members of the evaluation team. Who is responsible for producing which documents can be a part of the pre-evaluation orientation session.

## ***Initial Scoring of the Instrument Based Only on the State Plan***

It has been suggested that each team member should preliminarily score Section 1 of the instrument – based exclusively on a reading of the state plan – prior to arriving on site for the evaluation visit. The benefits of this process are that each team member will

have become thoroughly familiar with the state plan and with the evaluation instrument before arriving for the site visit. The drawbacks are that the team members will need to be willing to revisit their scores to reflect the additional information obtained during the visit. The plan itself will not include complete information about the process by which it was developed. The team will need to give the state credit for actions they have taken that are not included within the plan.

On balance, the Design Team believes this would be a worthwhile process improvement to implement. If this scoring takes place at least a month in advance of the site visit, it would provide the team with much greater understanding on which to base its requests for demonstrations or for focused interviews.

### ***Travel in the Course of the Onsite Evaluation***

Both the Washington and Kentucky evaluation teams objected to the amount of travel involved in the evaluation process. They recognized that travel produced good will on the part of the legal services programs themselves; the gesture of traveling to meet them was very much appreciated. But team members concluded that they did not learn very much by traveling around the state and visiting program offices.

The consultant benefited greatly from the drive from Spokane to Wenatchee and then from Wenatchee to Seattle, during which two members of the team were accompanied by one of the senior members of the Washington legal services community. The long drive provided almost eight hours of time for the three of them to interact. If travel were arranged so that evaluation team members were accompanied by senior legal services personnel, the travel time could serve as opportunities for extended discussions well beyond the limited time and scope available during the interview process.

The Ohio pilot was designed to test the effectiveness of a process without travel – where all of the persons to be interviewed would travel to a central location and the evaluation team would not travel. The team members reported that they believed they obtained the information they needed to score the instrument from telephone interviews.

The Design Team recommends that LSC take a “strategic” approach to including travel within a state evaluation site visit. The evaluation team should not travel around the state merely “to show the LSC flag.” However, they should be prepared to travel when it furthers an objective of the evaluation site visit. Examples of “strategic” travel include:

- travel to view a demonstration of a program unique to a particular program;
- travel to rural areas of a state with a mixture of rural and urban programs;

- travel to other urban areas when a state has competing centers of urban influence;
- travel to a state with markedly different geographic areas to appreciate concerns about geographic diversity;
- travel to visit individuals who have played an importance role in development of the state's legal services program;
- travel to address other issues of respect and courtesy, or to ensure the credibility of the evaluation.

The Design Team suggests that central interviews be held at a bar association, court, or legal services facility rather than in the sterile setting of a hotel room.

### ***Length of the Evaluation Site Visit***

The first two pilot tests were six days in length. The Ohio test was limited to three and a half days. The first two evaluation teams were comfortable with a week long data gathering process, being skeptical that they could have gathered sufficient information in a shorter period of time. The Ohio team did not have sufficient time to complete scoring of the instrument.

The Design Team suggests that there be no standard length for an evaluation site visit. Though it believes that most evaluations will need a full week, some small states will not require that much time and larger states will require more than a week. The LSC and state principal contacts will need to agree upon a time schedule as the interview, demonstration, and travel plans develop.

The Design Team suggests that each schedule guarantee the evaluation team sufficient time to complete the individual and consensus scoring processes on site. That will mean a half day for individual scoring and a full day for consensus scoring.

The Design Team suggests the following to save some on site time:

- As recommended above, have all team members complete their individual scoring of the first section of the instrument well prior to the site visit.
- Eliminate the opening session with the state leadership cadre. State leaders can afford to take time for one such meeting during a week; their attendance at the exit conference is far more important than an entrance conference. The pre-evaluation orientation session will serve most of the purposes of the opening session conducted during the pilot tests.

- Have team members conduct telephone interviews prior to arriving in the state for the site visit. There is no advantage to be gained by making such calls while within the state boundaries.
- Try to have the entire evaluation process completed within six weeks from the time of the first telephone interviews to the delivery of the written report to the state leadership team.

### ***Composition of the Evaluation Team***

The inter-rater reliability analysis concludes that the instrument should only be used with a group evaluation process. How large should the evaluation teams be? The pilot test experience suggests a basic team size of four members, increased for particularly large or complicated state evaluations. The Ohio team was not big enough. The Kentucky team was larger than it needed to be.

How should such groups be composed? The answer to that question is affected by budget and staffing realities as well as by methodological considerations. And the methodological considerations involve tradeoffs as well. The involvement of legal services peers and consultants – as in the Washington and Kentucky pilot tests – brings a breadth of practical experience and perspectives about the delivery of legal services to the judgments required to score the instrument. Restricting the teams to LSC staff members – as in the Ohio pilot test – might however ultimately lead to greater consistency in the scoring of the instrument through the accumulation of experience with the instrument, with the evaluation process, and with the variations in state planning processes from state to state in a small group of professionals within the LSC State Planning Team. These decisions appear most appropriate for LSC to make as it develops its experience with the evaluation process.

LSC has expressed the desire to involve legal services program directors and senior staff in future evaluations, to the extent that LSC budgets will allow. The Design Team suggested that LSC not compensate program directors participating in future evaluations; they should perform this service as volunteers, not as paid consultants. The Design Team urges LSC to ensure that future teams include a good mix of different skills and experience, such as someone with automation savvy, someone with experience with rural legal services program issues (for states with significant rural areas), and someone with the skills to facilitate the consensus scoring process and keep the group on task and on time.

### ***Maintaining a Record of the Evaluation***

The LSC principal contact should maintain a list and central archive of all written materials gathered in the course of the evaluation. S/he should also collect and maintain the interview questionnaires administered during the site visit. These materials should be

kept for one year and then discarded unless some controversy persists concerning a particular evaluation.

### ***The Access Database***

Following the Kentucky pilot experience, LSC and Greacen Associates agreed not to use the Access database for the Ohio pilot test. The experience in Kentucky suggested strongly that members of evaluation teams are more comfortable using printed copies of the instrument in the individual and team consensus scoring. Further, the use of the electronic scoring process did not save time in Kentucky; it used up time that the team needed for consensus scoring. Use of the Access data base requires that team members find personnel in their offices or communities to load the application onto their laptops; the evaluation team will need to include a member with sufficient automation capability to download the individual scores from each of the team member's laptops, and run and print out reports useful for the consensus scoring effort. This process also requires that each team be equipped with a portable printer.

The Excel spreadsheet attached to this report provides LSC with the capability to record consensus scores in electronic form for the generation of reports and analyses. Given the unreliability of individual scores for the instrument, it would not be appropriate for LSC to retain them in permanent electronic form.

Given these problems with the use of the Access database encountered during the pilot tests, and the availability of an alternative electronic format for maintaining and analyzing the final consensus scores, the Design Team recommends that any further development or refinement of the Access database be put on hold.

### ***The Format and Content of the Final Report***

Both Washington and Kentucky were critical of some of the observations made by the evaluation team in their narrative comments and prescriptive recommendations appended to the report. However, they both had generally positive comments on the report as a whole and particularly with the observations and recommendations made by the team members. Narrative summaries and recommendations for improvements should be included in each evaluation report.

While it is not necessary to report the scores to the state's legal services leadership during the exit conference, it is necessary to report the team's overall observations, the strengths and weaknesses of the state's justice communities planning and implementation process, and the team's recommendations for the most needed improvements in those processes. These observations need to be rooted in the team's consensus scores. Consequently, the evaluation team should complete the scoring before developing the exit conference report. The experience in Washington and Kentucky was that the narrative summary arose spontaneously from the scoring results. Neither team required more than two hours to agree on the themes for the exit conference, assign

responsibilities for presenting them, and preparing a PowerPoint presentation to support the presentation.

On the whole, Greacen Associates recommends that formal evaluations contain narrative overall observations and recommendations. This is in keeping with the general directions within the instrument itself that evaluators identify areas most in need of improvement.

### ***Use of the Interview Protocol***

LSC receives several complaints that pilot team members were not asking all of the interviewees all of the questions contained on the interview protocol. Persons expressing this concern were worried that the team members were not obtaining comprehensive information because of this practice.

The Design Team recommended that evaluation team members ask a limited number of standard questions and that the interview protocol be revised to reflect this requirement. The protocol should contain the standard questions, with instructions to ask additional questions related to the particular expertise and experience of the person being interviewed.

A revised protocol incorporating these suggestions – and the input of members of the three pilot evaluation teams – has been provided to LSC.

## **Issues With Respect to the Instrument**

### ***Recognizing the Collective Nature of the Evaluation Process***

The term “evaluator” has been replaced with “evaluators” wherever it appears in the evaluation instrument.

### ***Giving Credit for Activities That Are Occurring But Are Not Included in the Plan***

All three of the pilot tests proceeded on the principle that a state should receive credit for activities being conducted as a result of statewide effort, whether or not they appeared in the state’s written plan. The focus was, in effect, on the state’s “planning process” rather than on the state’s “plan.” The Design Team discussed this issue at considerable length, agreeing ultimately that while it may seem illogical to address activities not included within the plan itself, the integrity of the evaluation process

requires it. First, a state plan is not continually updated; the planning process continues, however. States should be given credit for planning initiatives undertaken at the state level whether or not they have been incorporated into a written plan document. The consistency and accuracy of the scoring of the instrument also requires the inclusion of what a state is actually doing rather than what appears within the four corners of its written plan. LSC records would be inaccurate if they overlooked significant progress in some area just because it was not included in the current version of the plan. Finally, this information cannot be gleaned usefully from the implementation analysis of Section 2 of the instrument. The listing of action items is not researchable in the same fashion as the subsections of Section 1.

However, the Design Team recommended that the instrument make clear that only those activities being pursued on a statewide basis be recognized in the scoring. The isolated efforts of one program within a state do not constitute a statewide initiative, unless the program has been recognized as a statewide pilot and other programs have committed to implement the process if it proves successful. Further, such supplementary activities should be identified explicitly in the evaluators' comments and the evaluators should suggest that they be incorporated into the written plan as soon as possible.

These principles have been added to the introduction to the instrument.

### ***Allowing the Use of a Broader Range of Scores***

The Kentucky evaluation team was unable to agree on a score of 3 or 4 for Section 2. It agreed to settle on 3.5. Several team members were sometimes unable to choose between two scores in their individual scores, reporting, for example, "3 or 4." Allowing the use of midpoints between all scores would provide more flexibility.

The LSC State Planning Team recommends the addition of this flexibility, with the understanding that the evaluator will be required in every instance to justify and explain the use of midpoints.

The Design Team rejected this suggestion. Great effort has been expended by the Team and the consultant to make the instrument as clear as possible. Criteria have been articulated for almost every score. Introducing additional scores, without standards for using them, does not seem wise.

### ***Changing the Labeling of the Parts of the Instrument***

The LSC State Planning Team finds the use of the terms Sections, Parts and Subparts confusing (particularly the introduction of A, B and C in Section 3). It would prefer that the larger units be referred to as Parts and the subparts as Sections and Subsections.

The Design Team members also reported continuing confusion about the structure of the document. While it did not believe the State Planning Team approach would produce greater clarity, the Design Team recommended that the numbering for each subsection of the instrument include a new introductory number indicating the Section of the document. For instance, subsection 3.10.1 of Part 3 of Section 1 of the instrument would be labeled 1-3.10.1.

Upon further review, Greacen Associates has concluded that the State Planning Team's recommendation has merit. The final version of the instrument will refer to its three main components as Parts. Subcomponents of the parts are referred to as Sections and subsections. The use of A, B and C in Part 3 has been converted to 1, 2, and 3 so that Part's organization parallels the organization of the rest of the instrument.

### ***Clarifying the Language in the Instrument***

The Design Team reviewed the evaluation instrument in detail following the pilot tests and made numerous additional wording changes and clarifications. One section (3D1 on the costs of state planning) was removed. A new subsection was added to Section 1 (1-1.1.10) to provide an unscored section in which to record any information a state is able to and wants to provide concerning the costs encountered in state justice communities planning. Several of the other objective measures were significantly changed, as discussed further below. A number of definitions were included or enhanced.

### ***Separately Scoring the Plan and Its Implementation***

Greacen Associates is perplexed by the difficulty evaluators experienced in separating their opinions of a state's legal services performance from their assessments of its planning process. As noted above, their negative views of aspects of the state's legal services program produced lower scores for the state's plan than were appropriate. During the Kentucky scoring process, the distinction was reinforced by the consultant and the resulting effect was minimized. However, it significantly colored the Washington scoring results.

We do not believe that the instrument should be changed in any fashion; the Design Team shares this view. However, we have included an additional section in the introduction alerting users of the document to the phenomenon encountered universally during the pilot tests.

The burden will fall upon the LSC staff directly responsible for the use of the instrument to ensure that evaluation teams include persons with the experience needed to identify this phenomenon when it occurs and to ensure that it does not affect the scores given for Section 1 (now Part 1) in states in which it is used for official evaluations.

## **Revising the Method for Scoring Section (now Part) 2-Implementation**

The approach to evaluating the implementation of a state's plan contemplated in the seventh draft of the instrument proved unworkable in practice. That approach called for five separate scores for each action item in the plan. In addition to an overall score, the team was to assess the "effort," "agility," "vision," and "achievement" of the state in implementing every action item.

The Washington evaluation ultimately included 45 items. The Kentucky evaluation included over 80 items. To make five different ratings of every action item would require 400 separate judgments for the Kentucky evaluation and over 200 for the Washington evaluation.

The Kentucky team, consequently, decided to use only the overall score for each action item, following the analytical framework set out in the seventh draft of the instrument. That process proved satisfactory. The Design Team recommends that it be incorporated into the final version of the instrument, with a few additional refinements and clarifications.

## **Section (now Part) 3-The Objective Measures**

Both the Washington and Kentucky teams were able to compute scores for parts A1, A2, and A3 of Section 3. Greacen Associates and the LSC principal contact were able to compute scores for parts A4, A5, C1, and C2 for Kentucky. Greacen Associates, working with the Northwest Justice Project, was able to compute scores for those same parts of Section 3 for Washington. Scores were not computed for B1 nor D1 for either state. Each of the measures is discussed below.

### **A1 through A3-Statewide Capacities**

These measures were relatively easy to use. The results for Washington and Kentucky are shown below. Note that the Kentucky score for A2 was misreported in the report to Kentucky. No changes are recommended, although the Design Team split one of the capacities in A2 into two separate capacities, increasing the highest possible score for that measure to 33.

	A1-Capacities to Improve Client Representation	A2-Capacities to Strengthen the Legal Services Community	A3-Capacities for enhancing public support and resource development
Washington	20/24 = 83%	20/30 = 67%	7/9 = 78%

Kentucky	10/24 = 42%	13/30 = 43%	5/9 = 56%
----------	-------------	-------------	-----------

These three measures clearly differentiate the more mature Washington statewide legal services structure from that in Kentucky, which has begun to emerge more recently.

***A4-Non LSC Resources Received by All Legal Services Providers Serving Persons Eligible for LSC-funded Services***

This measure proved relatively easy to use, with the exception of the requirement that states use 125% of poverty in reporting the number of poor persons in the state. Neither state was able to provide that number; both reported data based on 100% of poverty.

Washington had difficulty deciding which sources of funding should be included and excluded from the two measures. LSC staff should anticipate the need to provide authoritative interpretations of the terms in the instrument for particular factual situations, and to maintain a record of such interpretations over time for consistency.

The results for Washington and Kentucky are shown below.

	Dollars and cents from state and local government/per poor person generated during the last calendar year	Dollars and cents from the private sector/per poor person generated during the last calendar year
Washington	\$7.89	\$1.19
Kentucky	\$5.25	\$.27

No changes are recommended for this measure. If current census data is not available from which to calculate state poverty populations at 125% of the federal poverty guidelines, states should use data from the prior census, with the most recent updated estimates.

***A5-Relative Availability of Legal Services Lawyers***

This measure proved relatively easy to use, with the exception of the requirement that states use 125% of poverty in reporting the number of poor persons in the state. Neither state was able to provide that number; both reported data based on 100% of poverty. The results for Washington and Kentucky are shown below.

	Poor persons potentially eligible for LSC-funded legal services per FTE lawyer
Washington	7,378
Kentucky	7,243

It is interesting that Kentucky has slightly more lawyers available per poor person, despite Washington’s decided funding advantage (as shown in A4 above).

The Design Team recommends that this measure be expanded to include a second category of poor persons in a state per “case handler.” The definition of case handler will include attorneys, but also include paralegals who are providing legal services and not just intake eligibility determinations. That change has been incorporated into the final instrument.

***B1-Quantity of Service Provided by Type of Service***

This measure proved much too broad and unspecific. It calls for reporting all annual case closings and services provided by type of service provided, as defined by LSC’s CSR and MSR reports. The evaluation teams were provided with copies of the most recent CSR and MSR reports for both Washington and Kentucky. They include literally hundreds of separate data points. It would be infeasible to attempt to set them forth in the evaluation report. It is also not clear what purpose would be served by doing so.

The Design Team’s purpose in defining this measure was to have an objective measure of the state’s output of services to its clients. Perhaps that objective could be met by aggregating the CSR and MSR data into a few categories – such as total cases closed and services provided, total advice and counsel and brief services cases closed, and total extended representation cases closed. These numbers should probably be weighted by dividing them by the poverty population. The results for Washington –using CSR columns A and B for closed advice and brief services and CSR columns F, G, H, I and K for extended services – would have been:

	Closed cases per 1000 poor persons	Closed extended representation cases per 1000 poor persons	Closed advice and brief services cases per 1000 poor persons	Services provided per 1000 poor persons
Washington	16,917/612.370 27.6	698/612.370 1.1	10,927/612.370 17.8	154,977/612.370 253.1

While these are rudimentary, summary numbers, they would serve as a baseline for a state’s overall productivity at a point in time. The data for closed cases is available for past years, so some trend data could be developed for even the first evaluation of a state’s planning process. MSR data would not be available for past years.

The Design Team approved this approach, with one change. Advice and counsel and brief services cases will include CSR columns A through E. Extended representation will include columns F through K. Consequently the totals of the second and third columns will equal the first column.

### ***C1-Geographic Equity in Resource Distribution***

Both Washington and Kentucky were able to compute this measure as specified in the instrument. The results are set forth below.

	Discrepancy in services provided in most and least served counties
Washington	87%
Kentucky	98%

The issue with this measure is whether it is the most appropriate way to show geographic disparity in the delivery of services. The measure as created compares the most served county with the least served county, based on the number of extended representation cases arising from a county divided by its poverty population. This compares the two extreme values in the state. It does not necessarily constitute a representative measure of disparity in services delivered. It is theoretically possible, but highly unlikely, that there would only be one underserved and one overserved county in the state. The resulting score would dramatically overstate the disparity for the state as a whole.

The Design Team recommends that LSC use a more sophisticated statistic for this measure – the number of standard deviations from the mean for the closed cases per poor person per county data. This measure will require someone to enter all of the data for the state into a statistical package and compute a value for the standard deviation within the data. It will have to be done at LSC for each state. The state would not be able to compute the value for itself (absent the help of a college professor or a social science graduate student).

The Design Team also asked that a second measure be added computing the same statistics for all closed cases. Focusing only on extended representation cases skews the data toward one sort of legal service delivery. Supplementing this measure with all cases closed will provide more complete information and more accurately reflect the results of all legal services activities, including statewide hotlines.

These changes have been incorporated in the final instrument.

### ***C2-Equity in the Provision of Services to Groups of Clients***

This measure requires a state to provide data on the disparity of services based on race/ethnicity, age, gender and two additional categories of the state's choosing.

Washington reported its breakdown of extended representation cases by race/ethnicity, age and gender; it did not provide the corresponding breakdown of those groups within the state's poverty population. Washington did provide the population data for its two additional categories. However, it found the wording of this measure very

frustrating. It reported that the instrument’s definitions are “poorly expressed and confusing.” It provided five different calculations for each of its two chosen categories – Native Americans and female heads of households. It provided the ratio of the population of poor persons within those groups compared to the total poverty population, and the number of extended service cases for each of the groups, but did not divide the number of extended service cases for each group by the total number of extended service cases closed during the calendar year. The correct figures are as reported below:

Washington group service disparity

	Proportion of the poverty population	Proportion of extended representation cases closed
Native Americans	44,040/612,370 7%	40/698 6%
Female head of household	48,358/612,370 8%	157/698 22%

This is useful information – that Washington’s legal services programs are providing services to the state’s Native American population roughly proportionate to its share of the poverty population and services to its female head of household population greatly exceeding its share of the poverty population. That disproportionate service may be warranted by the special needs of the population group; but, this is useful information for Washington to have. It contrasts dramatically with similar data from Kentucky.

Kentucky provided race, age, and gender information, though it combined the race and age information in the same set of tables. It included one additional group – female heads of households. The Kentucky data, reported as intended, is set forth below:

Kentucky group service disparity

	Proportion of the poverty population	Proportion of extended representation cases closed
Asian	.6%	.2%
Black	12.6%	9.9%
Hispanic	2.1%	.5%
Native American	.4%	.3%
White	84.3%	86.2%
Other	0%	2.8%
Age under 18	32.4%	3.5%
Age 18 – 59	56.7%	90.5%
Age 60+	10.9%	6.0%
Male	44%	15.7%
Female	56%	84.3%
Female head of household	33.1%	35.6%

This information is even more interesting – suggesting possible under service of Blacks and Hispanics, persons under 18 and over 60, and males. Females and middle-aged poor persons may be over served, compared to their proportion of the population,

but female heads of households are not. (That anomaly suggests that Kentucky should look into how the staff of the four legal services programs record female heads of household data. The numbers of cases in which legal services provide extended representation to persons of “other” races and ethnic groups also suggest the need to review carefully how staff use the “other” racial category in reporting case information.)

The dramatic difference between the percentage of female heads of households in the Washington and Kentucky poverty populations (6% versus 33%) suggest that we should explore further the census reports used by the two states to derive their statistics.

Greacen Associates concludes that this measure is an important one and should be retained, but that states should be given a worksheet showing how to make the relevant calculations. The measure should explicitly state that differences between the proportion of closed extended representation cases involving persons of a particular group and that group’s proportion of the poverty population does not necessarily mean that the group is under or over served. It may mean that the group’s members have legal problems either in numbers or severity out of proportion to their representation within the poverty population.

The Design Team concurs in this recommendation, suggesting however that the requirement for additional population subgroups be reduced from “two” to “one or two,” that states be encouraged to identify population groups for which census data is unavailable, such as developmentally disabled persons, runaways, or homeless persons, and that the measure be expanded to apply to all closed cases as well as extended representation cases.

These changes have been incorporated in the final instrument.

### ***D1-Annual Costs of State Planning***

In response to comments from the field, the Design Team included a measure of the costs of state planning. Both Washington and Kentucky reported that they were unable to develop realistic estimates of the annual costs of the planning process, especially in the absence of a much more precise definition that differentiated development, revision, updating, and monitoring of the state plan from implementation of the plan and general efforts to improve the delivery of legal services within the state.

The Design Team recommended that this measure be dropped from the instrument. It recommended, instead, a new subsection 1-1.1.10 in which a state would have an opportunity to report the costs that it has incurred in conducting state planning. However, the instructions will note that the involvement of non-legal services volunteers in the planning process does not constitute a cost, but rather a contribution to legal services – in most cases as contribution that would not otherwise be made.

## **Additional Questions Identified in the Preparation Guide**

The Design Team identified a series of questions in the Preparation Guide on which it requested specific feedback from the evaluation teams and the pilot states. These questions resulted from the last Design Team meeting prior to the pilot tests. These items were not discussed specifically with the state leadership cadres nor necessarily addressed by the evaluation team members during their post-site visit discussions with LSC staff. Greacen Associates' comments on each follow the question.

Whether the items included in Part 1 of Section 1 should be scored or merely reported as information items. No comments received suggest that these topics be unscored. The LSC state planning team was afraid that it would not be able to obtain information on the visibility and stature of the state plan during a short site visit. The pilot tests did not encounter problems obtaining this information from interviews or in scoring these items. Some of the definitions for scoring have been clarified.

Whether the items in Part 1 of Section 1 should be moved or otherwise revised to reduce their visibility and apparent importance. Issues of structure did not dominate the evaluation process nor discussions with state legal services leaders. There is no need to reorganize the structure of the instrument.

The length of the instrument. Kentucky legal services leaders reported that they did not find the instrument too long when they read it in detail. They found its thoroughness helpful. The pilot tests did not disclose any major sections that proved superfluous.

The burden imposed on the state of providing the information for and participating in the evaluation process. Both Washington and Kentucky were asked "what sort of burden did the evaluation process create for the programs of your state?" Washington reported:

Surprisingly little. Yes, it was a major time commitment for those of us involved in the planning and coordination of the LSC team, but the visit itself was minimally intrusive. We appreciated the approach taken by team members on site and those back in DC.

Kentucky's response was "minor."

For LSC-funded programs the impact was minimal. Kentucky was fortunate to be able to dedicate three state level staff to the preparation for the evaluation and the on-site logistics, so that the programs only needed to comply with data requests. For this reason, we believe that states that do not have a state support office or dedicated state planning staff will find the process more burdensome at the program level – at least as the evaluation was conducted in Kentucky.

It is clear that states must devote resources to the evaluation process. But at least for states with state level staffing the burden of the evaluation process is not great.

Availability of data called for by the instrument. With the exception of data on the costs of the state planning process, both Washington and Kentucky provided the data called for, except for poverty populations at 125% of the federal poverty level. However, it is clear that the directions for presentation of the data were not sufficiently clear. They have been refined in the final version of the instrument.

Whether the use of 125% of poverty as the measure of the population of persons eligible for LSC services unduly complicates the data reporting process. Neither Washington nor Kentucky were able to provide 125% of poverty data. However, both states assured the Design Team that the data was not available from the Census Bureau at the time of the pilot test; it has since become available and could be provided with no effort today. The Design Team recommended that states provide information based on prior census reports, with update projections, until new census data becomes available. The LSC Office of Information will provide technical support to states needing help with census or other data; the evaluation coordinator from the State Planning Team will coordinate requests for such assistance.

Whether the exclusion of data from non-LSC-funded programs skews the data provided in the measures reported in Section 3. No comments were received on this issue, suggesting that participants in the pilot tests did not view the measures as flawed because they lacked data from non-LSC-funded programs. The Design Team revisited this issue. It asked that additional commentary be added to some of the objective measures, but did not recommend changing the measures.

Whether the definition of client-centeredness in Part 1.1.1 of Section 1 is sufficient. This definition did elicit comments and the Design Team made further refinements to it.

Whether Part 1.1.3 of Section 1 works as drafted. One of the pilot evaluation teams asked that the criteria be redrafted to be more parallel and clearly distinguishable from each other. Recommended changes were approved by the Design Team.

Whether Part 1.1.4 of Section 1 works as drafted. No comments were received on this issue, suggesting that participants in the pilot tests did not encounter problems.

Whether Part 1.1.8 of Section 1 works as drafted. No comments were received on this issue, suggesting that participants in the pilot tests did not encounter problems.

Whether Part 1.1.9 of Section 1 works as drafted. No comments were received on this issue, suggesting that participants in the pilot tests did not encounter problems.

Whether Parts 3.3 and 3.4 of Section 1 work as drafted and whether they should be combined into a single section. No comments were received on this issue, suggesting that participants in the pilot tests did not encounter problems.

Whether the preparation of legal materials for clients outside of a litigation context are appropriately reflected in the CSRs. No comments were received on this issue, suggesting that participants in the pilot tests did not encounter problems. The measure using CSR data has been revised substantially, but no one on the Design Team saw a need to supplement the CSR data itself.

Whether the instrument needs to limit evaluation of implementation of the state plan to only the highest priority actions. Both the Washington and Kentucky plans produced very long lists of implementation action items – demonstrating that aggressive state planning is ongoing in both states. While the idea of limiting the evaluation to the highest priority action items is attractive, it is not readily apparent how such prioritizing could be done or by whom. If it were done by the state being evaluated, it could skew the evaluation. It is not clear how the team could prioritize the action items in advance of the site visit, nor how prioritizing them during the scoring process would save any time or effort.

Whether paralegals should be included in the A5 measure in Section 3. Kentucky noted that the exclusion of paralegals produced incomplete data for this measure. The Design Team recommended that the availability of “case handlers” – including paralegals -- be added to this measure. The final instrument includes both categories for this measure.

Whether the instrument needs to be revised to provide a different means for the evaluator to provide feedback to the state. The instrument itself appears to serve as an appropriate vehicle for reporting the results of the evaluation.

The success of the team process, including the sharing of information, the number of persons needed to conduct interviews to gain a sufficient understanding of the state’s planning process, and the team’s ability to reach consensus on scoring of the instrument. These items are covered in the description of the pilot tests in Washington, Kentucky and Ohio included above.

LSC should be alert to the contents of Ohio’s response to the report from its pilot test to identify any additional issues that should be addressed in the instrument or in the process by which it will be administered.