Major Issues for Consideration in Revising The Program Assessment Rating Tool (PART): Preliminary Recommendations by the Performance Evaluation Team (PET)

Through OMB Spring Reviews and consultations with agencies and others, OMB's Performance Evaluation Team (PET) has received valuable feedback on the Program Assessment Rating Tool (PART). The PET is preparing to revise the PART and have a final version (with guidance) released in early July. This memo outlines six preliminary recommendations by the PET for resolving major issues that have been raised regarding the PART.

In general, based on testing of the PART in Spring Review and feedback from various sources, we believe that the PART is a useful assessment tool. The PART has helped to: formalize and make more consistent analyses that examiners typically perform; focus attention on program strengths and weaknesses; renew attention on establishing meaningful performance measures; and prompt agencies to provide more useful assessment data than they have previously been willing to provide. RMOs and agencies have also raised concerns about the PART; however, we believe these issues can be addressed within the existing structure of the PART. **We recommend using a revised version of the PART to rate programs and inform FY 2004 budget decisions.**

This memo presents major issues in two categories:

- Issues that could be addressed through clarified guidance, but would not require changes to the PART. These issues are presented in the first section below with a description of how we intend to handle in revised guidance.

- Issues that could require changes to the PART questions, format or scoring. We have presented these issues with options and recommendations for resolution.

ISSUES REQUIRING CLARIFICATION BUT NOT CHANGES TO THE PART

Clarification of the PART Role: OMB staff must have a clear, consistent, unified view of what the PART is and is not. Many OMB staff concerns reflect a lack of common understanding of how the PART results will be used to inform budget decisions. RMOs have correctly noted that the PART does not frame budget and policy trade-offs for senior officials, demonstrate how changes in funding affect performance, or automatically determine a specific funding level for a program. RMOs and agencies are uneasy because performing a PART analysis layers more work on top of existing work, and have requested additional information on how this effort will fit with other for Fall Review.

It will be important for OMB policy officials to reinforce a consistent view of the PART's role. Specifically that: 1) the PART will be one of many pieces of information that will go into budget decisions, 2) it will focus agency attention on strategic planning, meaningful performance measures, and program results, 3) it will inform budget decisions by helping to diagnose the source of program success or failure and by providing directional information on how a program is performing, and 4) it will not, by itself, determine program funding levels.

Resolving Differences in Rating within OMB and between OMB and Agencies. Both the PMC and RMOs have asked for further clarification on the process for resolving differences between OMB and agency assessments. In addition, we have noticed that within OMB, RMO approaches to rating similar programs differed.

The PET recommends that agencies and RMOs would be expected to work together to resolve any differences in ratings, but ultimately the PADs would resolve PART issues, just as they do on many issues and conflicts between OMB and agency staff. The PET will clarify this relationship in guidance, including creating a schedule for when PART scores need to be finalized.

An alternative to PAD-centered resolution of PART score disputes would be to create a centralized OMB-wide appeal process. The PET does not believe this is practical, given the number of programs to be rated, nor advisable. To the extent possible, PART scores should be handled just like other conflicts that arise during a budget process between agencies and OMB. If they are treated differently, they will not truly be integrated into the traditional budget decision-making process.

Relationship of the PART to GPRA. RMOs expressed a lot of confusion about the relationship between the existing GPRA measures and processes and the PART. As the Director noted, the PART helps focus and apply GPRA principles more directly to budget decisions. Specifically, it helps OMB work with agencies to establish good program performance measures, not merely ones for which good data exists. We plan to clarify the PART guidance to reflect this view and change the PART format to require listing of a program's key performance measures. We will also need a clear OMB statement for public use on the relationship of the PART to GPRA.

ISSUES FOR DECISION

Decision Issue 1: **Should there be a Program Purpose/Federal Role section in the PART? If there is a Federal role section, should that section be more "objective"? Should it capture whether the program is a presidential priority or initiative?**

Some believe the Federal Role section should be weighted more heavily since it addresses whether a program is necessary at all. Initially, some had the opposite view, suggesting this section be eliminated, since it seems less objective than the rest of the PART and addresses issues beyond the control of the program. Hill staff have also suggested this. On a related issue, several policy officials expressed that a program should be assessed on its own merits, regardless of whether it is a priority for the President, while others argued that programs that are Presidential priorities should automatically be assessed as having a critical Federal role. The issue came up specifically in the discussion of Americorps – an ineffective program, but a probable vehicle for the President's high priority volunteerism objectives.

NOTE: These options are independent of a decision on how to factor program purpose into the overall score.

Option 1: No significant change in the PART, but refine the questions and guidance to more clearly address program design issues. Change name of the section to "Program Purpose and Design" – no reference to "Federal Role."

Option 2: Eliminate the Federal Role section.

Option 3: Retain the Federal Role section, and add a question on whether a program is a presidential priority or initiative.

PET Recommendation: **Option 1**

Decision Issue 2: **Should the composite score be retained and displayed in the budget?**

One of the most consistent comments made during Spring Review is that the individual PART section scores are more informative than a single composite PART score number. Some examiners recommended we avoid a composite numerical score (e.g. 63%) since this implies false precision. Others felt strongly that a composite score focused agency attention on improving program management and performance, and creates greater incentive to provide data.

On the related issue of section weighting, some suggested that the PART section weights be shifted, but there is no consensus on how to do this (some wanted a higher weight for results, others for management, and others for Federal role). Still others questioned whether the RMOs should be allowed to change the PART section weights to address program-specific issues, e.g. to avoid penalizing new programs that have yet to produce results. We will likely be revisiting this issue after the reviews are completed.

There have also been several comments about how section I (program purpose) should be factored into the overall score. Some think it should be shown separately, but not used in calculating the score. Others think it should be used as a filter or threshold test – if a program scores poorly on program purpose, it should be rated ineffective, no matter how well the other sections score.

There are also a number of ways to present scores in the Budget. For example, numeric scores could be banded and translated into letter grades. This would avoid arguments over whether a program with a 65 is really better than a program with a 62. Below are a eight options for presenting scores in the Budget, with brief explanations of differences. Note that in these options we have added a "minimally effective" rating as a gradation between "moderately effective" and "ineffective." Samples showing how each of these options could be presented in the budget are on the following page.

Option 1: Leave scoring and section weighting as is. In the Budget, display composite and section scores. This approach has the advantages of consistency across programs, simplicity of overall assessment presentation, and provides clear motivation for agencies to improve their ratings.

Option 2: Eliminate the composite PART score and only present the section scores. This approach has the advantage of presenting a more informative assessment than a composite score and providing a somewhat clearer indication of why certain recommendations were made. However, we would lose the "shorthand" presentation of assessment results, as well as a valuable incentive for agencies to implement program improvements in order to receive a better PART score. A summary score is still calculated and used to arrive at the overall rating, although it is not presented.

Option 3: Present the composite numeric score, but separate out program purpose so it does not calculate into the overall score. In the example below the weights are as follows: 20% for Strategic Planning, 20% for Program Management, and 60% for Results.

Option 4: Program purpose section score is used as filter or threshold. For example, if you score 50% or less on program purpose, the program rating is automatically ineffective. If program purpose were above 50%, the overall score and rating would be calculated the same as in option 3.

Options 5 through 8 are essentially the same as options 1 through 4, except that letter grades replace numeric scores in the budget presentation. Note that numeric scores would still be used to calculate the overall score and arrive at the letter grades, but they would not be presented in the budget. However, it is likely that the underlying numeric scores would still be available publicly if the PART worksheets are posted on the internet.

For the purpose of these examples, the following bands for letter grades were used (we would need to agree on these bands): 86 – 100 = A; 71 – 85 = B; 56 – 70 = C; 41 – 55 = D; 0 – 40 = F.

Option 5: Essentially the same as option 1, with letter grades.

Option 6: Similar to option 2 - summary/composite score is not presented.

Option 7: Similar to option 3 – program purpose is presented separately and not used to calculate the overall score.

Option 8: Similar to option 4 – if program purpose is a D or E, then the program is automatically ineffective. If purpose is C or above, then scores would be calculated without considering the program purpose section score.

PET Recommendation: **Option 5**

Sample Presentations of Scoring Options:

Option 1: Present Composite Numeric Score/No Change in Weighting (as in Spring Review)						
Department of Commerce	Program	Strategic	Program		Overall	Effectiveness
Program:	Purpose	Planning	Mgmt.	Results	Score	Rating
Economic Development Asst.	45%	100%	80%	60%	**65%**	**Minimally Effective**

Option 2: Present Sections Score Only						
Department of Commerce	Program	Strategic	Program			Effectiveness
Program:	Purpose	Planning	Mgmt.	Results		Rating
Economic Development Assist.	45%	100%	80%	60%		**Minimally Effective**

Option 3: Present Composite Numeric Score/Purpose is separated out.						
		Program Effectiveness Rating				
Department of Commerce	Program	Strategic	Program		Overall	Effectiveness
Program:	Purpose	Planning	Mgmt.	Results	Score	Rating
Economic Development Assist.	45%	100%	80%	60%	**72%**	**Moderately Effective**

Option 4: Composite Numeric Score/Purpose is a filter (e.g., less than 50% automatically ineffective)						
		Program Effectiveness Rating				
Department of Commerce	Program	Strategic	Program		Overall	Effectiveness
Program:	Purpose	Planning	Mgmt.	Results	Score	Rating
Economic Development Assist.	45%	100%	80%	60%	72%	**Ineffective**

Option 5: Present Composite Letter Grade/No Change in Weighting (as in Spring Review)						
Department of Commerce	Program	Strategic	Program		Overall	Effectiveness
Program:	Purpose	Planning	Mgmt.	Results	Score	Rating
Economic Development Assist.	D	A	B	C	**C**	**Minimally Effective**

Option 6: Present Sections Grades Only						
Department of Commerce	Program	Strategic	Program			Effectiveness
Program:	Purpose	Planning	Mgmt.	Results		Rating
Economic Development Assist.	D	A	B	C		**Minimally Effective**

Option 7: Composite Letter Grade/Purpose is separated out.						
		Program Effectiveness Rating				
Department of Commerce	Program	Strategic	Program		Overall	Effectiveness
Program:	Purpose	Planning	Mgmt.	Results	Score	Rating
Economic Development Assist.	D	A	B	C	**B**	**Moderately Effective**

Option 8: Present Letter Grades/Purpose is a filter (e.g., D or less is automatically ineffective)						
Department of Commerce	Program	Strategic	Program		Overall	Effectiveness
Program:	Purpose	Planning	Mgmt.	Results	Score	Rating
Economic Development Assist.	D	A	B	C	**D**	**Ineffective**

Decision Issue 3: **Should banded scores and assignment of overall effectiveness rating be relative (i.e., on a curve), or absolute?**

Related to the scoring options in Issue #3 above, is the question of scaling of overall scores. There are two basic options:

Option 1: Rate programs on a curve assuming that program scores fall out in a bell-shaped distribution. This option would likely appeal to agencies who think the PART standards are too high and who suspect that the process is rigged to rate a lot of programs ineffective. Under this option, the bands of scores would not be able to be determined until PART scores are final (in October).

Option 2: Use an absolute scale based on high standards of program performance. For example, under #1 (rating on a curve) an overall score of 60 could be "moderately effective," but under #2 (absolute scale), it might be "minimally effective" or "ineffective." This option is consistent with the ratings for PMA scorecard. For example, getting a "green" on management initiatives is an absolute, high standard.

Recommendation: **Option 2**

Decision Issue 4**: Should the yes/no format be changed?**

Many RMOs have raised concerns about the "Yes/No" answer for each question, but staff are divided on whether to switch to a multi-point scale. Those supporting a scale would like to give some credit for partial achievement. They believe that the "Yes/No" answer sets too high a standard for success in some areas. For such situations, they have recommended that the question permit an answer that would show such partial compliance, e.g., responding on a scale from 0-3 to indicate a tendency toward Yes or No, rather than a full Yes or No. A 4-point scale would also help capture partial achievement of goals or uneven performance by program sub-components. Others strongly support the clarity of "Yes/No" answers and believe that approach helps focus agency attention on key program assessment issues. They have found it particularly helpful in the management section.

Option 1: Continue to permit only "Yes/No" answers and make the standard of evidence clearer in the guidance.

Option 2: Use a mixed approach. Include a 4-point scale in the Results section, but retain "yes/no" in section 1-3. This would help capture partial achievement of program goals, but retain the clarity of yes/no answers in the diagnostic sections of the PART.

Option 3: Use a 4-point scale in all sections to increase RMO flexibility.

Recommendation: **Option 2.**

Decision Issue 5**: Should new programs be assessed differently from established programs?**

Several RMOs have highlighted the difficulty of evaluating new programs, particularly because program results are not available for several years.

Option 1: No change in PART. Expectations will need to be adjusted to expect low ratings for new programs for the first few years.

Option 2: Score new programs on Sections 1-3 only, excluding Results, perhaps for the first 2-3 years (the timeframe would need to be agreed upon in advance). The advantage of this option is the PART can still be applied to new programs and provide useful information on agency planning and management efforts. This may require displaying new and established programs separately in the Budget.

Option 3: Score new programs for all sections, but include questions in Results that give credit to strategies that have a strong chance of resulting in an effective program. The disadvantage with this option is that the PART will be based less on documented evidence and will be left open to criticism for applying even more subjective judgment.

Recommendation**: Option 2.**

***Decision Issue 6*: Giving Credit for Program Improvements.**

Some RMOs have pointed out that the PART penalizes agencies for programs that had problems in the past but that the agency is on track to correct. This raises the issue of whether the PART should give credit for plans before evidence indicates improved results. A few of the NAPA panel participants also suggested giving credit for plans for program improvement to provide an incentive for agencies to implement changes.

Option 1: No change in the PART. Clarify guidance to indicate how to use strategic planning and management questions more effectively to capture plans for program improvement. This keeps a consistent focus on results, but could be discouraging to agencies that are implementing rigorous efforts to improve performance.

Options 2: Modify the strategic planning or management sections of the PART to ask specific questions about improving programs. This would give more credit for improvement plans, but could create some overlap in the questions in those two sections. Using a 4-point scale in the Results section would address the related issue of getting partial credit for partial results.

Recommendation: **Option 2**