



**USAID**  
FROM THE AMERICAN PEOPLE

## **Recommendations for Improving Methodology of Monitoring Implementation of the Justice Sector Reform Strategy in BiH**

Date: January 2011

This publication is submitted for review to the US Agency for International Development.

Prepared by: Andrew Green

## Introduction

The current methodology for assessing progress on the JSRS Action Plan is problematic, and as a result is not as useful for monitoring Action Plan implementation as it could be. Assessment of the current “traffic light method” (TLM) finds it to be an unreliable measure, which moreover may not be measuring what in the reform process is of interest to judicial system experts, international organizations, donor agencies, knowledgeable civil society leaders, or citizens. The assessment also finds that the later process of verifying statuses is also problematic. Recommended changes include adopting a problem-based logic for activity status assignment, developing explicit status assignment guidance, adding a ‘priority’ flag, revising verification processes, and enhancing the semi-annual progress report.

## Analysis of current TLM

Methodologists focus on two key aspects of data quality: reliability and validity. The concept of *reliability* is based on reproducibility, i.e., one person’s measurement of something could be re-created by someone else. Issues include stable and consistent data gathering procedures, as well as the consistent application of measurement or scoring criteria. The concept of *validity* is based on representativeness, i.e., it corresponds to or is an accurate reflection of what was intended to be measured. Issues include the translation of theory into practice, the range of conditions covered by the measure, and the purpose of the measure.

Of the two key aspects, validity is perhaps the most problematic for JSRS implementation monitoring, although reliability is also severely flawed. As a result of both, the semi-annual progress report cannot serve the important information role as originally envisioned.

### *Validity*

***Monitoring of Action Plan implementation needs to be better connected to the purpose of monitoring and even the nature of the activities.*** Validity is not just about measuring what is intended to be measured, but also whether what was intended to be measured is appropriate or useful. In the case of the current TLM, the current logic of ‘degree of progress’ is different than the original JSRS logic of ‘delays to be resolved by different actors.’ Both are about the pace of JSRS implementation, but each provides different insights into the process. Which is more useful for monitoring JSRS? Interviews with FWG members strongly demonstrated that the progress-based logic generated arguments and in the end provided little useful information for understanding Action Plan implementation.

There are four reasons why the current progress-based logic is flawed and therefore inappropriate:

- The statuses of ‘Not Completed,’ ‘Partly Completed,’ and ‘Fully Completed’ are based on activity progress, but do not in reality represent that. The ‘Not Completed’ category includes activities that are not even planned to begin until after the current reporting period. Debates over whether an activity is ‘Partly’ or ‘Fully Completed’ demonstrate that those terms may not represent what people think they do.
- Not all activities are based on the same amount of effort or complexity, i.e., not all activities are created equal. A progress-based logic treats each activity the same, though, so that an analytical report activity has the same weight as a new law activity.
- Not all progress toward activity completion is the same, either. Some activities take a short time to complete; others are multi-year efforts by nature. Clearly, the amount of progress that the latter could achieve in six months is likely to be far less than the former.
- There is no systematic way to flag delayed activities for resolution at any level, nor is it possible to ascertain how serious the problem or delay is. Similarly, there is no way to identify activities as

higher priority. The current progress-based logic simply lumps together delayed and priority activities in the same “Partly Completed” status with everything else.

### *Reliability*

**Monitoring of Action Plan implementation unquestionably needs clearer and more objective guidance for assessment as well as verification.** Discussions with FWG members and monitoring report consumers quickly uncovered a significant reliability problem: inconsistent application by the FWGs of the “Partly Completed” or “Fully Completed” statuses, stemming as much from a lack of clear guidance as from the nature of some activities.

Lack of clear guidance on status assignment is a key problem for reliability: if FWG assessors do not know what criteria to apply, then each assessment exercise will apply different criteria. Clear, objective criteria are fundamental to reliability, and hence to quality measurements.

Reliability problems are particularly acute for subjective matters such as classification of ongoing activities, e.g., implementation of a plan. Is such an activity “Fully Completed” because the plan was approved for implementation, or is it “Partly Completed” because the plan has not been fully implemented? Another example would be when one justice sector institution has completed its task, but the others have not. That activity would be “Fully Completed” from the perspective of the one institution, but “Not Completed” or “Partly Completed” from the others’.

There will always be subjectivity in assessing processes, it is unavoidable. The key is to constrain or limit how subjectivity is exercised, to provide objective rules to guide subjective assessments, so that someone else could reach the same subjective conclusion. The current TLM provides no means for constraining subjectivity, other than a review (or ‘verification’) of FWG assessments that is, itself, not constrained.

### *Reporting*

The validity and reliability problems discussed above have an impact on the semi-annual progress report, too. Reporting on the status of activities across the Action Plan based on the current TLM logic presents a misleading and unhelpful picture. Activities that have not yet begun are labelled “Not Completed,” contributing to the number and size of the status. “Partly Completed” is an exceptional large status that includes activities that are progressing without problems, experiencing minor delays, or have stopped almost completely pending action by an institution outside the judicial sector. Even “Fully Completed” is potentially misleading, because re-classification of an ongoing process that experiences an unexpected problem would result in *fewer* activities in this status.

The use of percentages as a measure of overall progress is also problematic, as it creates a false sense of precision. Calculating 64.83% of activities with a particular status is not meaningful when activities are treated as a whole – not all activities are of similar effort or duration, and not all progress is similar.

## **Improving the TLM**

The validity, reliability, and reporting of JSRS Action Plan implementation can be improved with relatively few yet key revisions to the current practice.

### *Strengthening Validity*

There are three main revisions to implementation assessment that would strengthen validity. First, adopt a problem-based logic instead of the current progress-based logic. Given that activities and progress have considerable variation in meaning, and that the chairpersons and members of the

Technical Secretariat for Monitoring the Implementation of the Action Plan (TS) and FWGs can obtain and understand details about any particular activity's progress, the implementation assessment approach should focus on identifying *delays* in progress, as something of an 'early warning' system. This is very similar to what was originally intended.

Second, the 'traffic light' scheme should be retained, as it is well-suited to qualitative, subjective measures. Moreover, numbers are not very meaningful, e.g., a score of 2 on a 5-point scale or a rating of 20% complete is meaningless when effort, duration, and progress vary significantly from activity to activity.

Third, priority activities should be identified and marked as such, preferably on a binary basis. Again, not all activities are created equal, and some activities are much more important to Action Plan implementation than others. There needs to be a way to recognize and focus on those activities.

### *Strengthening Reliability*

Two revisions would strengthen reliability. First and foremost, develop explicit guidance for assigning a status to an activity. Since activities are actually aggregations of many individual tasks to be performed by particular actors over a period of time, the guidance should be based on whether an activity as a whole is being delayed by even one task, and then on the severity of that delay. Second, verification of all assigned statuses should be performed using the same guidance, the same comments, and the same documentation that was used by the FWG. In the absence of detailed plans and timelines for tasks within an activity, it is obvious that there will still be some subjectivity in status assessment; however, subjectivity has been constrained to a judgment about when a problem should be resolved, which would also be reviewed in the verification process.

### *Enhancing Reporting*

The semi-annual progress report as it currently exists provides vast amounts of information for experts deeply involved in particular parts of the reform process. For those not involved in particular parts of the process, or not involved directly at all, the report is far less useful. Revising the TLM should generate more useful information. The semi-annual report should provide information on the status for all activities, priority activities, and pillar activities, as well as analysis of any patterns to be found in the set of activities experiencing significant delays.

## **Specific Recommended Revisions to the TLM**

Recommended steps for improving the TLM are aimed at strengthening the validity and reliability of the assessments, as well as on verification and reporting:

1. Adopt a modified form of the monitoring and evaluation scheme outlined in Section 8 of the JSRS, which was a problem-based logic instead of the current progress-based logic.
2. Retain the 'traffic light' scheme.
3. Add a 'high priority' flag to a reasonably small number of activities.
4. Develop explicit criteria to guide FWGs in their assessment of activities.
5. Revise the current process of verifying FWG-assessed statuses.
6. Strengthen the analytical and communication qualities of the semi-annual progress report.

The end result should be JSRS activity assessments that are more valid and reliable, and a report that is more meaningful for reform actors, international organizations, NGOs, and citizens. Each step is discussed in more detail below.

### *Validity*

1. **Adopt a modified form of the monitoring and evaluation scheme outlined in Section 8 of the JSRS, which was a problem-based logic instead of the current progress-based logic.**
2. **Retain the ‘traffic light’ scheme.**
3. **Add a ‘high priority’ flag to a reasonably small number of activities.**

The current TLM is based on logic of ‘degree of progress,’ although the JSRS called for logic of ‘delays to be resolved by different actors.’ A modified version of the original logic would produce assessments that are much more useful to domestic and international reform actors, as well as to consumers of the progress reports. A revised TLM would essentially function to provide an early warning of problems as they emerge during implementation, rather than letting them publicly explode later.

The chairpersons and members of TS and FWGs are the main people who need to know and understand the details of progress for any specific activity, but all actors directly or indirectly involved in the JSRS Action Plan efforts have a much stronger need to know about implementation problems generally. Those who need details of activity progress would consult available information on the new documentation system, while others primarily need information at the pillar or overall strategic reform level.

The ‘traffic light’ scheme should be retained, not least because it provides an easy and visual picture of implementation progress. Such schemes are much better suited to categorizing qualitative information. Numeric schemes would weaken validity and introduce considerable unreliability, particularly if the underlying activities do not have the same level of effort or complexity, duration and deadlines are not the same, and progress has different meanings. Such schemes tend to produce not just arguments over the exact number to be assigned, but also lead to a false sense of precision.

It is recommended that the following ‘traffic light’ signals be used generally as follows:

- **GREEN** – The activity is experiencing no or only transient delays
- **YELLOW** – The activity is experiencing minor delays that need attention
- **RED** – The activity is experiencing significant delays that need attention

Note that this differs slightly from the original logic of the JSRS Action Plan document, in that it focuses on the severity of delay, not on which authority level should be notified. The rationale for this is that not all large problems need the attention of the MC, just as not all small problems can be resolved by a FWG. What is important is *how significant the delay* is. In order to avoid trading one subjective criteria, ‘degree of completion,’ for another, ‘severity of problem,’ explicit guidelines must be established so that FWGs are able to assess problem severity in a clear and justified manner (see next section).

The assessment scheme has two criteria, how long it will take to resolve a delay in a task within an activity, and whether resolution depends on an institution other than the activity’s lead institution(s). Two assumptions are at the foundation of these criteria: first, the longer it takes to resolve a delay, the bigger the problem is; second, the more that delay resolution depends on another institution, the more difficult it will be to coordinate resolution efforts. Note that severity should also be greater when an activity is a priority. The table below lays out the logic and criteria in visual form:

### Standard

		Across Institutions	
		No	Yes
<b>Time to Resolve Task Delays</b>	None, all tasks on schedule	<b>G(reen)</b>	<b>G</b>
	Before next FWG meeting	<b>G</b>	<b>Y(ellow)</b>
	Before next MC meeting	<b>Y</b>	<b>R(ed)</b>
	After next MC meeting	<b>R</b>	<b>R</b>

### Priority

		Across Institutions	
		No	Yes
	None, all tasks on schedule	<b>G</b>	<b>G</b>
	Before next FWG meeting	<b>Y</b>	<b>Y</b>
	Before next MC meeting	<b>R</b>	<b>R</b>
	After next MC meeting	<b>R</b>	<b>R</b>

Some activities are more critical or high-profile than others. The Ministerial Conference should mark a reasonably small number of activities as priority, upon the recommendation of participating institutions and the FWGs.

## Reliability

### 4. Develop explicit criteria to guide FWGs in their assessment of activities.

The assessment guidance first identifies activities that are truly complete and those that are not scheduled to have begun, which leaves activities that are in-progress. A following question identifies priority in order to direct the assessment to specific questions and decision flowcharts. The questions guide the assessment through whether an ongoing task within an activity is on schedule, a delay in that ongoing task will be resolved before the next FWG quarterly meeting, a delay in that ongoing task will be resolved before the next MC semi-annual meeting, or a delay in that ongoing task will be resolved after the next MC semi-annual meeting, as well as whether the delay must be resolved by an institution other than the lead institution(s) for that activity. Note that the deadline being referenced here is not the activity deadline, but a 'working' deadline set by the FWG for a task within an activity; this is analogous to how chemical reaction rates depend on the slowest reaction.

1. Is this activity scheduled to begin prior to the next FWG quarterly meeting?
    - If yes, then go to Question #2
    - If no, then the status is **NOT STARTED**

---

  2. Is this activity complete, i.e., have all tasks been performed and the FWG has no plans to work on this activity again?
    - If yes, then the status is **COMPLETE**
    - If no, then go to Question #3

---

  3. Is this a priority activity?
    - If yes, then go to Question #9 and use the priority activities decision flowchart
    - If no, then go to Question #4 and use the standard activities decision flowchart

---

  4. Are all ongoing tasks within the activity being performed according to deadlines set by the FWG for those tasks?
    - If yes, then status is **GREEN**
    - If no, then go to Question #5
  5. Is there an ongoing task within the activity that will miss its deadline, but be completed by the time of the next FWG quarterly meeting?
    - If yes, then go to Question #6
    - If no, then go to Question #7
  6. Does responsibility for the overdue task lie with another institution?
    - If yes, then status is **YELLOW**
    - If no, then status is **GREEN**
  7. Will that overdue task be completed by the time of the next MC semi-annual meeting?
    - If yes, then go to Question #8
    - If no, then status is **RED**
  8. Does responsibility for the overdue task lie with another institution?
    - If yes, then status is **RED**
    - If no, then status is **YELLOW**
-

9. Are all ongoing tasks within the activity being performed according to deadlines set by the FWG for those tasks?
- If yes, then status is **GREEN**
  - If no, then go to Question #10
10. Is there an ongoing task within the activity that will miss its deadline, but be completed by the time of the next FWG quarterly meeting?
- If yes, then status is **YELLOW**
  - If no, then status is **RED**

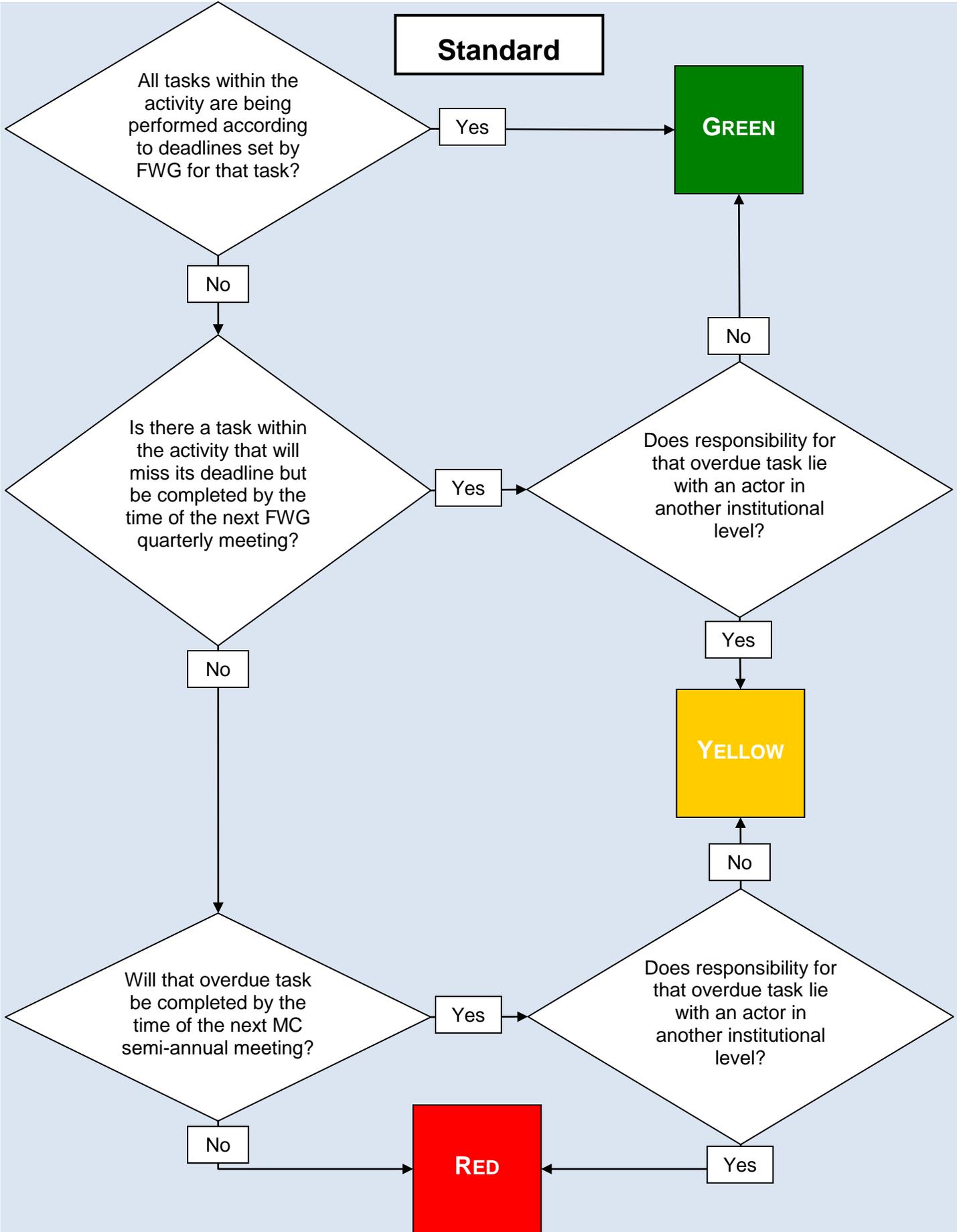
### *Verification*

The verification process should be conducted using the same guidelines and information (documents, comments) as the FWG. It should be noted that the proposed criterion of the 'working' deadline for a task within an activity is open to some subjective judgment; however, one purpose of verification is to examine how accurate such judgments are.

### *Reporting*

The semi-annual progress report should be a source of extensive information for any reader. The report should first disaggregate reporting at the overall and pillar levels according to the binary priority marker. A simple overall chart should be produced that shows completed as well as activities progressing with no delays. The report should also account for activities that have not yet started, as well as activities that have been completed. It should also add analysis of the most serious delay status, identifying institutional levels that are not performing as well as needed.

# Standard



**Priority**

