Introduction

There are currently 16 UN peacekeeping missions deployed in 15 countries around the world. With a combined budget of $7.3 billion and 115 troop- and police-contributing countries deploying nearly 90,000 uniformed personnel alongside staff from the Departments of Peacekeeping Operations and Field Support, UN peacekeeping is the largest and costliest part of the organization’s operations worldwide.

While substantial, the current level of operations represents a substantial decline from 2010, with 10,000 fewer troops and $200 million cut from the annual budget. This decrease results in part from an ongoing effort to review and reconceptualize peacekeeping operations, in light of the evolving challenges to international peace and security, not to mention the growing financial burden that the expansion of peacekeeping has placed on the organization. Extended deployments and “mission creep” have led policymakers and diplomats to push for greater clarity and narrowed focus in terms of peacekeeping mandates and objectives.

One of the central challenges facing these efforts is how to determine when a peacekeeping mission has accomplished its objectives. A broad consensus among member states and within the Secretariat emerged that past UNSC mandates have been unclear in this respect. Furthermore, missions in post-conflict situations struggle to prepare the relevant institutions for peacekeepers’ departure, culminating in an unsuccessful transition from peacekeeping to peacebuilding efforts. These issues have featured prominently in the conversation surrounding the effectiveness of peacekeeping in general. In the wake of the global financial crisis, the issue of funding has also become a central point. Balancing the
desires for more effective peacekeeping alongside the need for member states to cut budgets has lead to a reexamination of the UNSC’s mandating procedures.

The benchmarking process is one result of these efforts which has been increasingly mainstreamed by the UNSC over the past several years. This paper investigates the UNSC’s practice of requesting UN peacekeeping missions to establish explicit benchmarks for their own progress and eventual drawdown. It provides a brief background, investigates the varying approaches to benchmarks across missions, and evaluates some of the emerging trends in this relatively new process. Each mission’s approach is discussed individually in an Annex following the paper.

**Development of the benchmark generating process**

The UNSC's demand for benchmarks reflects a growing difficulty in evaluating progress, as the traditional cease-fire monitoring of the Cold War era gives way to multidimensional missions with broad and complex mandates focused on internal conflict. As the 2008 Capstone Doctrine notes, “Determining whether a multidimensional United Nations peacekeeping [operation] deployed in the aftermath of a violent internal conflict has successfully completed its mandate is far more challenging given the number of complex variables involved.” This difficulty, combined with the dramatic growth of UN peacekeeping operations in the past decade, has left the UNSC with the prospect of large, resource-intensive missions with no exit strategy or prospects for withdrawal. Asking missions to detail explicit benchmarks for their own exits thus reflects UNSC members’ desire that peacekeepers either find a way to accomplish their mission, wrap up and come home, or acknowledge their limitations and withdraw.

The UNSC has requested benchmarks for the majority of the UN’s sixteen peacekeeping missions, beginning with Liberia 2007 and expanding the practice in 2009. These benchmarks are subsequently developed by the mission and the Secretariat, and eventually presented to the UNSC in an Annex to the Secretary-General’s regular report on the mission’s progress. The UNSC typically endorses the benchmarks in its subsequent renewal. UN missions in Darfur, southern Sudan (and subsequently South Sudan), Côte d’Ivoire, Chad and CAR (now withdrawn), Liberia, Afghanistan, Haiti, and Timor-Leste have provided benchmarks. No benchmarks have been requested for the new UN stabilization mission in Mali.

**Benchmark structure**

*Core versus Contextual*

UN peacekeeping operations’ fundamental objectives are established in the mandate provided by the UNSC. However, UNSC resolutions often touch on factors outside of the mission’s direct purview—peacekeepers frequently operate under mandates which include detailed objectives they are unable to accomplish directly. For example, resolutions often

---

6 While the UNSC has the authority to reject or modify the mission’s proposals, in practice the body has typically endorsed benchmarks without comment, occasionally placing a stronger emphasis in certain areas.
call on governments, rebels, or regional powers to make political accommodations, engage in peace processes, or otherwise support resolution of conflict. While a peacekeeping mission can create the environment for such progress to occur, it is not always expected or able to bring the parties to the table.

In order to evaluate a peacekeeping mission's progress in implementing its mandate, peacekeepers' core responsibilities must be distinguished from contextual objectives. This distinction is one of the key reasons for establishing benchmarks. Distinguishing between core and contextual benchmarks should also help the UNSC identify how to address a lack of progress. If core benchmarks are not being met, the mission may require reconfiguration or reorientation. On the other hand, if contextual benchmarks are not being met, other actors (member states, international or regional organizations) may need to address obstacles in other political or economic forums.

Several peacekeeping missions have used benchmarks to successfully delineate between their core and contextual objectives. The UN mission in Liberia (UNMIL) was one of the first missions mandated to produce benchmarks for drawdown, and explicitly defines its core versus contextual benchmarks, explicitly indicating which benchmarks fall into which category. The UN Assistance Mission in Afghanistan (UNAMA) also takes some pains to clarify the mission's responsibilities in each benchmarked area (see the annex for more detail).

However, delineation of core and contextual benchmarks is often neglected where it is most critically important: when missions are mandated to “assist” or “support” provisions of peace agreements, leaving ambiguity as to the scope and depth of their involvement. Benchmarks provide an opportunity to clarify this vagueness and specify the mission’s “core” responsibilities, but often fall short. Both the UN missions in Côte d’Ivoire (UNOCI) and southern Sudan (UNMIS), for example, are mandated to support peace accords. In both cases, however, the benchmarks fail to clarify for what aspects of the agreement the mission is directly responsible. Without clearly delineated responsibilities, the benchmarks lose much of their utility as metrics for evaluation of progress, as the UNSC cannot readily identify the source of an obstacle or shortcoming.

---

7 “Missions should distinguish the core benchmarks for which peacekeepers are responsible from broader targets which reflect wider progress in peace consolidation and rely on the performance of others.” A New Partnership Agenda: Charting a New Horizon for UN Peacekeeping (July 2009), p. 15.
10 The UN mission in Côte d’Ivoire (UNOCI) is mandated to “assist” the Government with restoration of state authority, and the mission has articulated a corresponding benchmark. However, the indicator “Deployment of mixed police and gendarmerie units in the north” does not clarify responsibility. If it remains unfulfilled, is this a failing of the mission, the parties to the Ouagadougou Peace Agreement, or international donors? UNMIS’ benchmarks have similar issues; see the annex for a full breakdown of these two missions.
11 Most missions do take pains to specify “assumptions” or “requirements of progress,” external factors that, if altered or unmaintained will inhibit progress in the mission’s mandate (continued donor support and cooperation of the host country are two common examples). While related to “contextual” benchmarks, by definition outside of peacekeepers’ “core responsibilities” and
**Indicators of progress**

Benchmarks are usually associated with indicators, specific tasks or steps to be taken which demonstrate progress against the more broadly-defined objectives of the peacekeeping mission’s UNSC mandate. For example, the UN mission in Chad/CAR (MINURCAT)’s benchmark, “capacity of local authorities to provide the necessary security for refugees, internally displaced persons, civilians and humanitarian workers,” is associated with a number of indicators with regard to building Chadian capacity, e.g. “Establishment by DIS of a security coordination mechanism with the Chadian National Police, the Gendarmerie and the Garde nationale et nomade.”

The exact form that these indicators take varies from mission to mission. UNAMA goes so far as to provide every indicator with quantifiable “metrics for evaluation,” often to be evaluated by external actors. This practice can increase the mission’s accountability: by relinquishing control of evaluation metrics, missions tie their own hands when reporting their progress to the UNSC. Such explicit evaluation metrics are not typical, but where they are absent, concrete and specific indicators of progress can provide similar accountability. A well-defined, clear and concrete indicator will be more straightforward to evaluate, even if the means of evaluation are not specified.

By contrast, missions that do not establish indicators for progress damage the utility of their benchmarks. The UN Integrated Mission in Timor-Leste (UNMIT) outlines benchmarks such as “The respective roles of the various institutions of the security sector are clearly defined,” and articulates no indicators for progress. This benchmark is sufficiently vague as to make meaningful evaluation difficult, breaking the link between benchmarks and substantive progress in achieving the mission’s objectives. In this case, for example, if security roles are defined but the East Timorese military maintains key responsibilities which should fall to the National Police, the benchmark may technically be met, even while the underlying conditions for instability remain. Clear indicators are essential to ensure that a desire to demonstrate progress does not supplant real efforts to improve the situation.

**Common benchmarks across mandates**

**Protection of civilians**

While the UNSC has articulated increasingly ambitious peacekeeping mandates for the protection of civilians, troop-contributing countries have pushed back hard at the robust rules of engagement required to realistically implement the concept. The retraction of

certainly useful while evaluating a mission’s progress, listing these requirements does not resolve the larger problem of ambiguous responsibility regarding the benchmarks themselves.

14 Naturally, this is not foolproof – see page 5 on electoral benchmarks.
peacekeeping operations from its peak in 2010 has further hampered efforts to establish meaningful civilian protection. Despite much discussion, there is no consensus as to what exactly protection of civilians entails from one situation to another, or even whether doctrinal consensus is useful given the complex and diverse context under which peacekeepers operate.

Protection of civilians is thus an interesting point of analysis for the benchmark system, as missions navigate between the UNSC mandates and the contributors they rely on, attempting to articulate strategies that please both sides. While the political space is crowded, the doctrinal vacuum has left room for benchmarks to influence or bypass the ongoing debates in New York, providing test concepts and working methods from the perspective of the peacekeeping missions themselves.

As in other areas, benchmarks range from concrete to vague. MINURCAT provides three clear benchmarks for “security and protection of civilians.” While some indicators leave room for interpretation, most are explicit, detailing numbers of DIS officers to be trained, security infrastructure to be developed and tasks to be completed. Thanks to their specificity and reliance on measurable indicators, these provide one of the clearest pictures of a protection mandate in the UN peacekeeping system. The benchmarks also played a central role in discussions following the Chadian call for MINURCAT’s withdrawal in 2010, when officials in the UN Secretariat made reference to protection benchmarks in their arguments that objectives remained unmet and that drawdown was premature.

By contrast, while protection of civilians was a core aspect of the UN Mission in Sudan (UNMIS)’s mandate (distinct from its mandate to support the CPA), its security benchmark included only two indicators of progress on protection which fail to tie benchmarks to concrete steps or specify how to evaluate progress (see Annex). Fulfillment of UNMIS’ indicators for its security benchmark would not necessarily demonstrate any actual progress in protection of civilians. Its utility as an evaluation tool is thus quite limited. In contrast, the benchmarks for its successor mission in South Sudan (UNMISS) establish a number concrete and measurable indicators for the protection of civilians, and focus on building the Government’s capacity—but leave responsibility for their attainment unspecified.

Electoral support

The UN’s controversial role in the 2009 presidential elections in Afghanistan and the 2010 presidential elections in Côte d’Ivoire underline the importance of explicitly defining the role of peacekeeping missions in supporting elections. Clear delineation of responsibilities are key, as a mandate from the UNSC to “support” elections can be interpreted quite broadly. Unfortunately, instead of clarifying the situation most electoral benchmarks are

17 The Office for the Coordination of Humanitarian Affairs (OCHA) presented a strong argument to the UNSC for MINURCAT’s continuation, noting that of the three protection benchmarks (endorsed by the UNSC), two remained unfulfilled and one had just begun to see progress. While the UNSC ultimately withdrew the mission in the face of Chadian intransigence, the protection benchmarks may have played a role in the UNSC’s reluctance for a precipitous withdrawal. If nothing else, they performed their functioned as an evaluating tool.
dangerously vague in spelling out responsibilities, paving the way for future missteps down the road.

UNAMA provides a unique case to analyze, as its benchmarks were created a full month after elections, but before the final results were announced. Released in the midst of a developing crisis in Kabul and New York, the benchmarks tried to leave the UN room to maneuver. Benchmarks thus emphasized that elections be “acceptable” to the populace, avoiding questions of massive fraud. However, evaluation metrics linked to assessments by national and international observers ultimately restricted the ability of UNAMA to claim that elections had advanced the mission's mandate.

UNOCI provides a contrasting example, with an SRSG directly mandated to certify every step of the electoral process. The mission’s indicators (mirroring agreed-upon timetables) functioned in the pre-electoral period insofar as they informed the UNSC of the repeated failures to meet electoral milestones, and clearly established where responsibility for the delays would fall. Where they fell short is in failing to list contextual benchmarks or “requirements for progress,” e.g. continued commitment by all parties to implementation of the Ouagadougou Agreement. Without providing this context, benchmarks can lose their relevance as they are overtaken by events on the ground, as made dramatically clear by the resumption of conflict following the 2010 elections. However, even here benchmarks (and the mandate backing them) played a crucial role, providing justification for UNOCI's stance in opposition to then-President Laurent Gbagbo due to massive electoral irregularities.

Human rights and the rule of law

Missions take two quite different approaches to benchmarking in human rights. MINURCAT's indicators focused on DIS capacity, policy development to address child soldiers, a measurable increase in investigation and prosecution, justice-sector training, and eventual ratification of international covenants and compacts. By narrowing their focus, MINURCAT's benchmarks were clear and measurable, providing the mission with specific objectives and highlighting problem areas for the UNSC's attention.

UNAMA by contrast focuses its indicators on broader social change in the population's behavior and customs. “Improved impact of, and support for, measures (including legal and policy reform) to combat violence against women and girls” is a laudable goal. However, it is simultaneously incredibly ambitious for an external peacekeeping mission, and difficult to measure. Such indicators are more aspirational goals than concrete objectives. Given the ubiquitous and broad nature of human rights goals in UNSC mandates, well-defined benchmarks with concrete objectives and indicators for progress are essential. Aspirational objectives can establish a context within which to create benchmarks, but should not serve as requirements for withdrawal.

Evaluation and Recommendations

The benchmarking process is still relatively new. Most missions have only just moved from establishing benchmarks to regularly measuring progress against them. Nevertheless we can make some preliminary assessments of the varied evaluation structures and benchmark systems that missions have established.
Clear, specific and concrete: Achievable benchmarks

Despite the importance of clear and concrete mission objectives, ambiguity does have its role to play in defining the mandate of a peacekeeping mission. It allows UNSC members to reach consensus (or avoid vetoes) while establishing or renewing peacekeeping missions. There are also strategic reasons for providing peacekeepers with flexible operating constraints, allowing the mission to adjust its activities and configuration in response to evolving situations on the ground. NGOs and activist groups often criticize missions of using conservative reading of mandates as an excuse for inaction. While clarification of roles is important, the ability to adjust to circumstances on the ground is crucial.

However as noted earlier, benchmarks are articulated not by the UNSC, but by the mission itself. Benchmarks thus provide a golden opportunity for the UN mission to establish clarity of purpose in areas where political concerns may hinder the UNSC’s own action. Furthermore, benchmarks are intended to provide a map for the mission’s withdrawal. A lack of clarity or specificity does not merely lessen their usefulness as a tool for evaluating progress; overly-broad benchmarks could actually sabotage a mission’s withdrawal by creating unrealistic objectives and expectations. There is a significant difference between indicators for progress that require establishment of corruption commissions, training human rights defenders and enacting legislation, and indicators that describe a fundamental change in a population’s outlook or society. Particularly in the context of a larger effort to reign in deployment lengths, a poorly defined or overly broad benchmark could force the mission to either modify the benchmark, ignore it, or manipulate its own analysis to argue that the objective has been met, undermining the entire process. While there is space for ambitious objectives in the UN system, such processes occur over a longer time frame than that of a typical peacekeeping mission.

Supporting peace agreements: Explicit responsibilities

Missions mandated to support peace agreements already have their own set of benchmarks, after a fashion: the provisions of the agreement serve as an obvious tool for evaluating the mission’s progress. Nevertheless, a mandate to “support” or “assist” implementation of a peace agreement often fails to clarify where responsibility for key tasks lies. When benchmarks simply mirror peace agreements, they may create ambiguity regarding the breakdown of core and contextual objectives. If DDR programs are not advancing, for example, or an election is not taking place, the reasons for the lack of progress will not be clear from an evaluation of the benchmarks themselves. Naturally, benchmarks can never stand alone; context and analysis will always be necessary. Nevertheless the mission should at least attempt to articulate its core responsibilities in the context of the peace agreement, thus highlighting for UNSC members (and other multilateral or bilateral partners) when obstacles outside of the mission's purview require attention.

Timing: Establishing benchmarks early

As emphasized in the Capstone Doctrine, benchmarks should be established at the earliest stage possible in mission planning. When benchmarks are established... Reliable
media res by a mission which has been on the ground for a year or more, they are created in a political and military context which can severely restrict their utility. UNAMA’s electoral benchmarks are a perfect example; faced with the prospect of a serious electoral crisis, the mission provided objectives designed to leave the mission room to maneuver and deflect criticism. Rather than reflecting a strategic assessment of the conditions necessary for the mission’s withdrawal, the benchmarks attempted to legitimize the situation on the ground.

In contrast, the Secretariat first laid out benchmarks for MINURCAT in December 2008, months before the transfer of responsibility from EUFOR to the UN in May 2009. Established early, the benchmarks became a useful tool for evaluating the mission’s progress and provided ammunition for the Secretariat in the face of intransigence from the host country. While no benchmark or indicator of progress will ever be free of political considerations, early adoption (if benchmarks are well-conceived) can force strategic thinking at an early stage, serve as a reminder of the mission’s core objectives and increase accountability on all sides.\(^\text{19}\)

**Conclusion**

In the unfortunate reality of budgetary constraints and “fatigue” over longstanding missions, the original strategic and operational concerns which gave rise to the benchmark concept are almost afterthoughts. It is perhaps ironic in this context that the benchmarking system does demonstrate real potential as a tool to provide strategic coherence across the UN system, improve communication between the UNSC and missions, and increase accountability for all of the actors involved. Of course, these ideals will only be realized if benchmarks are well-conceived. To date, benchmarks and indicators have often been developed in missions which have been on the ground too long; most are fraught with vagueness and ambiguities that hinder their usefulness as evaluating mechanisms. As the practice becomes more mainstreamed, earlier and realistic benchmarking could address many of these issues.

Beyond the quality of the benchmarks themselves, the success of the system will depend significantly on two factors. One is the UNSC’s ability to follow up now that benchmarks have been worked into mission’s regular reporting. Without continued evaluation and review from the UNSC, benchmark generation runs the risk of becomes another box to tick for mission staff. Any number of initiatives have been launched since the Brahimi Report in an attempt to reform peacekeeping practices, but lack of UNSC follow up remains one of the primary reasons for such initiatives’ ultimate failure. That the UNSC has not yet requested DPKO to establish benchmarks for the new peacekeeping mission in Mali is troubling in this regard.

The second factor is the willingness of the mission and the Secretariat to articulate objectives which they may fail to meet. Faced with UNSC members trying to shorten

---

benchmarks and indicators are required to determine when the United Nations peacekeeping operation can begin the process of hand-over and withdrawal, without jeopardizing ongoing efforts to consolidate the peace.” United Nations Peacekeeping Operations: Principles and Guidelines (2008), section 10.2, page 88

\(^{19}\) Although its benchmarks functioned properly, MINURCAT was ultimately hamstrung by the lack of a political mandate, leaving too many of the benchmarks for withdrawal outside of its control. This serves as a reminder that benchmarks cannot substitute for a well-conceived political strategy.
deployments, mission planners will be tempted to articulate a more “flexible” roadmap: the Secretariat can interpret benchmarks broadly to declare success, UNSC members can announce that objectives have been achieved, and the mission can withdraw. However, the UNSC has not yet forgotten relapses in Haiti or Timor-Leste, where pressure for swift withdrawal left behind unstable situations and outbreaks of further violence forced the UN to return, or in Côte d’Ivoire, where long-delayed elections restarted the conflict that had been frozen since the signing of peace accords. A tightening budget has not entirely suppressed the desire to leave behind a sustainable peace, and so long as this remains true, benchmarks have the potential to provide a useful framework for UN peacekeeping operations.
Annex: Mission-Specific Analysis

MINURCAT

The SG first presents benchmarks for MINURCAT's withdrawal to the UNSC in his report on 4 December 2008. The benchmarks (developed in consultation with OCHA, UNDP, UNHCR, OHCHR and MINURCAT) cover six areas:

a) Voluntary return and resettlement in secure conditions of a critical mass of internally displaced persons;
b) Demilitarization of refugee and internally displaced person camps as evidenced by a decrease in arms, violence and human rights abuses;
c) Capacity of local authorities to provide the necessary security for refugees, internally displaced persons, civilians and humanitarian workers;
d) Ability of the national law enforcement agencies to maintain law and order with respect for international human rights standards;
e) Progress of an independent and effective judiciary in eastern Chad contributing to end impunity, as evidenced by a significant increase in the capacity and independence of, and respect for, the justice sector;
f) A strengthened prison system in eastern Chad that is based on a human rights approach to prison management.

The UNSC subsequently endorsed these benchmarks in resolution 1861 (2009) and asks the SG to develop a strategic work plan, stressing in particular:

a) Voluntary return and resettlement in secure and sustainable conditions of a critical mass of internally displaced persons;
b) Demilitarization of refugee and internally displaced person camps as evidenced by a decrease in arms, violence and human rights abuses;
c) Improvement in the capacity of Chadian authorities in eastern Chad, including national law enforcement agencies, the judiciary and the prison system to provide the necessary security for refugees, internally displaced persons, civilians and humanitarian workers with respect for international human rights standards.

Following this endorsement, the SG presents this workplan in an Annex to his 17 July 2009 report. In the report’s body, he notes that “The plan outlines several indicators and related activities for each of the benchmarks. It should be noted, however, that responsibility for meeting the benchmarks for withdrawal goes beyond the mandate and capacity of MINURCAT and the United Nations country team.” He particularly highlights the Government of Chad’s responsibilities in this regard. The report also adds a benchmark on “peace and security in the subregion.”

The Annex begins by listing “external” factors necessary for the fulfillment of benchmarks but outside of MINURCAT’s direct control or mandate (e.g. continued donor support for the DIS). The seven benchmarks are then sorted into three general areas: 1) security/protection.

---

20 This analysis does not incorporate the developments of May 2010, when UN/Chad negotiations resulted in significant modification of the mission’s mandate and a drawdown schedule for full withdrawal by December 2010.
of civilians; 2) human rights/rule of law; and 3) regional peace support. Each benchmark is articulated, and then broken down in detail with a timeline and “indicators of progress and activities.”

As an example, Benchmark 2 under Security/PoC is “demilitarization of camps for refugee and internally displaced persons as evidenced by a decrease in the circulation of arms and the level of violence and human rights abuses.” The ensuing timeline presents four indicators of progress. The first indicator, “enhancement of community policing in key towns, refugee camps and internally displaced person sites,” then has four activities listed, such as “850 DIS officers fully equipped and armed in all 19 DIS police stations and posts in eastern Chad and with demonstrated ability to maintain law and order in refugee camps and key towns.”

Granted, some indicators are vaguer (also under Benchmark 2, “Enhancement of wide-area security” is to be achieved by “Attainment of full operational capacity by the MINURCAT military component and provision of wide area security coverage throughout eastern Chad”). Furthermore, certain key indicators are not detailed as fully: “Improvement of security in the places of origin of internally displaced persons” has no corresponding activities listed. However, where provided, this hierarchical structure provides a clear picture of what tasks are planned in order to achieve progress against indicators and meet benchmarks. Furthermore, listing these tasks usually provides an implicit acknowledgment of who is responsible for progress in the area.

Overall, while responsibilities are not delineated as explicitly as in UNAMA’s benchmarks, the delineation of tasks (where provided) presents a clear picture and evaluating tool for progress in achieving the mission’s mandate and conditions for drawdown. This was particularly notable during negotiations over the government’s request that the mission be withdrawn. OCHA, for example, has cited these benchmarks and pointed to the specific uncompleted indicators and tasks – which the UNSC has endorsed – as an argument for renewing the mission. When activities as specific as “Establishment by DIS of a security coordination mechanism with the Chadian National Police, the Gendarmerie and the Garde nationale et nomade” remain outstanding (and scheduled to be completed by July 2010), the argument that MINURCAT be immediately withdrawn faces at least one extra hurdle.

MINUSTAH21

In Resolution 1780 (2007), the UNSC requested that the SG develop “a consolidation plan with appropriate benchmarks to measure and track progress, in consultation with the Haitian government.” The SG subsequently outlined five priority areas in his 22 August 2007 report:

1. Resolution of political differences through dialogue and completion of the election cycle;
2. Extension of State authority;
3. Establishment of an appropriate security structure;

21 This analysis here is based on the mandate and SG reports preceding the January 2010 earthquake, and has thus been overtaken by the humanitarian crisis. Nevertheless, is presented as an example of a possible approach to benchmarking.
4. Development of independent, credible and effective judicial and penal institutions;
5. Improvement in socio-economic development.

These became the basis for a more comprehensive consolidation plan, outlined in his 27 August report the following year. The report’s Annex defines a benchmark for each area, lists requirements for progress, and provides an extensive list of indicators for progress against this benchmark (with a suggested timetable).

Initially, the indicators for areas 1 and 2 are not very specific (e.g. “Continued collaborative approach in finding solutions Haiti’s political, economic and social problems”). However in the 2009 report, concrete measures are attached to these more vague indicators. Area 3 (security) gives quite specific indicators regarding police reform, and notes in several places where MINUSTAH support is expected. Area 4 indicators (judicial/RoL institutions) are similarly concrete and specific. Area 5 indicators (socioeconomic development) are vague, although the report explicitly acknowledges that the area is not covered by MINUSTAH’s mandate.

Whether responsibility is explicitly defined varies by benchmark. Sometimes, responsibility is contained within an indicator (“Updated legislative agenda is implemented collaboratively by the Government and Parliament”), but often not (“Preparation of a follow-on development plan to the current Haitian National Police reform plan”). This may be intentional, reflecting debates between UNSC members over the extent of MINUSTAH’s peacebuilding role. Overall, however, MINUSTAH’s role in a given area is made fairly clear in the list of requirements for progress, which typically note the need for continued support from MINUSTAH in given areas (notably absent from area 5, social and economic development).

Thus while peacekeepers’ core responsibilities are implied on the broader “area” level, there is still ambiguity regarding responsibilities from indicator to indicator. As suggested, this may be due to disagreements in the UNSC over what in fact the mission should handle (resulting in a vague mandate). Nevertheless, the benchmarks’ strength is in their concrete and specific indicators for progress. They are useful for evaluating progress, even if the mission’s responsibility is not clearly distinguished from the broader context of progress.

UNAMA

The UNSC requests the SG to develop benchmarks in resolution 1868, 23 March 2009:

36. Requests the Secretary-General to report to the Council every three months on developments in Afghanistan, and to develop, for inclusion in his next report, benchmarks for measuring and tracking progress in the implementation of UNAMA’s mandate and priorities as set out in paragraph 4 of this resolution, and calls on all actors concerned to cooperate with UNAMA in this process;

The SG lays out these benchmarks in Annex I to his report of 22 September 2009. The report provides a comprehensive description of the purpose and intent of its indicators, which cover the following areas: 1) Governance and institution-building; 2) Security; 3) Economic and social development; 4) Human rights; 5) Counter-narcotics.
The report is quite explicit in delineating who is responsible for what. In the Annex’s introduction, each area is briefly discussed, specifying UNAMA’s responsibilities in the context of the larger international effort. For example, area 2:

The second set of benchmarks addresses improvements in security. The key actors are Afghan and international security forces (except the benchmark on civil-military coordination, which comprises a specific element of the UNAMA mandate), although UNAMA has the capacity to help monitor progress towards these benchmarks.

Each section articulates the benchmark, notes “external” requirements necessary for progress, and provides indicators for progress in list form.

Unusually, the initial report spells out exactly how and by whom the benchmarks are to be evaluated, “metrics” for each indicator. For example, under area 4 (human rights), the indicator “Improved awareness by Afghans of their rights and by the Government of its obligations” is linked to the metric “Assessment based on monitoring by UNAMA and the Afghan Independent Human Rights Commission.” Some indicators, such as “Increase in the number of police and army personnel mentored, trained and operational according to an agreed structure” (area 2, security) have a great number of metrics that require action by various international partners.

The combination of these two clarifications—the initial delineation of responsibilities and explicit metrics for evaluating progress against benchmarks—make UNAMA’s benchmarks, on paper, very straightforward to measure, and they are at least theoretically well-defined. On the other hand, it is easy to pick these benchmarks apart in hindsight: “Holding credible national and subnational elections” has metrics including “Assessment based on UNAMA monitoring of the electoral process” and “Acceptability of results by voters, candidates and political parties.” As is now clear, these metrics for assessment leave plenty of room for ambiguity, confusion and even manipulation. Nevertheless, many of the benchmarks provide concrete indicators for progress, delineate responsibility, and provide metrics for measuring which (in some cases) seem to increase accountability for the mission.

**UNMIL**

UNMIL has had benchmarks for perhaps the longest of any UN mission. In 2007, an assessment mission provided recommendations for benchmarks linked to drawdown of the mission. Listed in the SG’s August 2007 report, these benchmarks were composed of six areas:

1. Completion of the basic training of 3,500 personnel of the Liberian National Police by July 2007
2. Completion of police operating procedures by December 2008
3. Completion of the formation of the 500-strong Police Quick Reaction Unit by July 2009
4. Equipping of police personnel and their deployment to the counties, as well as building of police infrastructure by December 2010
5. Finalization of the national security strategy and architecture and their implementation throughout the country by December 2008
6. Training and operationalization of the first and second Armed Forces of Liberia
The benchmarks, designed specifically to measure progress necessary for the mission’s drawdown, are (by the standards of other UN missions) incredibly focused. They do not cover broad topics such as human rights, complex political processes such as elections or ill-defined objectives such as protection of civilians.

While more limited in scope, the benchmarking process benefits as a result. The SG’s 19 March 2008 report details a number of more specific benchmarks within these areas, each with indicators of progress. These benchmarks explicitly distinguish between “core benchmarks” and “contextual benchmarks,” even defining the two terms in the Liberian context:

Core benchmarks: markers to measure progress in the achievement of the mandate and conditions needed to ensure a security “steady-state”; security “steady-state” is defined as the point at which the national security services and agencies are fully operational and primarily sustained with the Government of Liberia’s own resources...

Contextual benchmarks: conditions or factors that could potentially reignite violent conflict or interact with core benchmarks to undermine security.

This clear delineation of core vs. contextual is more explicit than in any other mission. Furthermore, in providing very specific benchmarks and still more specific indicators—and in defining some of the more potentially vague terms (such as “operational”)—the mission and UNSC are provided with remarkably clear and concrete evaluating tools. This is again more feasible because of UNMIL’s more focused mandate has (compared to larger missions such as MONUSCO) for supporting security-sector reform.

In fact, in distinguishing between core and contextual benchmarks so clearly, the benchmarks become a useful tool for directing international support. Where should donors put their efforts if they want to ensure the mission will fulfill its mandate? Presumably, areas covered by contextual benchmarks would be a good place to start.

**UNMIS**

In Resolution 1870 (2009), the UNSC “requests the Secretary-General to develop benchmarks for measuring and tracking progress in the implementation of UNMIS mandate.” The SG produces benchmarks in his following report (14 July 2009). As UNMIS was mandated to assist and support parties in the implementation of the CPA, “Implementation of its mandate is thus linked to the commitment of the parties to the Comprehensive Peace Agreement process and their progress achieved in implementing its provisions.” Benchmarks cover the following areas:

1. Referendums
2. Border demarcation
3. Abyei
4. Elections
5. Southern Kordofan and Blue Nile States
Each area initially contains several indicators. These range from specific events or objectives to be completed, to more general objectives. In their first incarnation, they are quite perfunctory and contain none of the context in which most missions place benchmarks (requirements for progress, etc).

The benchmarks are significantly revised in the SG’s following report (21 October), moving away from a laundry list of incomplete CPA provisions and towards the structure used by most other missions. The areas are restructured into:

1. Ensuring security and stability
2. Elections
3. Referendums
4. Implementation of CPA protocols
5. Wealth-sharing
6. Humanitarian recovery and development
7. Human rights and rule of law

For each area, the benchmark is defined, requirements for progress are listed, and indicators are provided (complete with vague timeline and, if applicable, date completed). Thanks to the nature of the UN mission in Sudan at the time, these benchmarks closely follow CPA provisions. However, no distinction is made between objectives with which UNMIS will assist, and objectives for the parties to complete without UNMIS assistance. This may be due to a desire not to simply parrot the long list of tasks laid out in Resolution 1870. However, UNMIS is mandated to carry out other tasks aside from the CPA, including protection of civilians “under imminent threat of harm.” While the Area 1 benchmarks do include reference to PoC, they are fairly weak in their metrics for progress—they do not require a measurable drop in violence, for example—and UNMIS’ role in PoC is not clarified.

Thus, while the benchmarks are useful in evaluating progress against the CPA, they are not particularly more useful than the CPA itself; one can fairly easily track progress against the timetables laid out in the back of the agreement. In fact, it seems that in focusing on CPA provisions the benchmarks miss a chance to clarify UNMIS’ other responsibilities with benchmarks.

UNMISS

Requested in Res. 1996 (2011), UNMISS' benchmarks are first outlined in the SG's 26 June 2012 report. With the CPA nominally complete and South Sudan independent, the focus of the benchmarks necessarily shifts from the pre-referendum UNMIS benchmarks, and are sorted into the following areas, each with a number of associated “indicators of achievement.”

1. The Republic of South Sudan has developed sufficient capacity to prevent, mitigate
and resolve conflicts and effectively carry out its responsibility to protect civilians.

2. The Government of the Republic of South Sudan has laid the foundations, at the national and local levels, for governing effectively and democratically.

3. The capacity of the Government of the Republic of South Sudan security and the rule of law institutions to maintain public order and administer justice is strengthened.


5. The Government of the Republic of South Sudan has laid a foundation for sustainable development in South Sudan.

For the most part, indicators are concrete and measurable, even where their scope is broad (for example, under Benchmark 1: “The incidence of large-scale armed violence in South Sudan is reduced consistently over time,” or under Benchmark 2: “By 2014, the new Constitution of the Republic of South Sudan is adopted and accepted as a basis for governance by all involved parties and it provides for women’s representation”). Some are more ambiguously defined or harder to measure (e.g. “Effective local and national mechanisms to mitigate and resolve conflicts have been developed, and the public perceives that the Government has this capacity”), but in general the indicators are concrete and in some cases quite specific.

However, while the benchmarks are mostly clear or concrete, they do not distinguish between “core” and “contextual” benchmarks, nor do they delineate responsibility between the government, UNMISS, and other stakeholders. Given the ambitious scope of the mandate, these distinctions are crucial when evaluating progress on the mission’s mandate, and for member states and the UN system to direct support or attention in the necessary directions. This is particularly the case since the benchmarks are articulated in terms of government capacity. Without clearly defined responsibilities, this creates the risk that UNMISS could be “held hostage” to capacity shortfalls which it cannot directly address.

UNOCI

The UNSC first officially requests benchmarks from the Secretary-General on UNOCI in Res. 1826 (2008). The benchmarks are for the specific purpose of reconfiguring the mission:

9. Expresses its intention to review by 31 January 2009 the mandates of UNOCI and the French forces which support it, as well as the level of troops of UNOCI, in the light of the progress achieved in the implementation of the key steps of the peace process and of the progress of the electoral process, and requests the Secretary-General to provide to it a report in this regard three weeks before this date, including some benchmarks for a possible phased drawdown of the troop levels of UNOCI, taking into consideration the electoral process and the situation on the ground and in particular the security conditions...

The Secretary-General responds in his January 2009 report, following a technical assessment mission. While proposing an initial drawdown, the report goes on to present benchmarks for next phases. These reflect the general contextual/core responsibility distinction – paragraph 47 provides general indicators of progress required for the mission’s drawdown, while paragraph 48 provides the mission’s “core responsibilities.”
47. The assessment mission also determined that the prevailing threats, the challenges ahead and the uncertain pace of the electoral process rendered it difficult to consider a more substantial drawdown of UNOCI's force until progress was made in carefully benchmarked areas and in security conditions on the ground. In that connection, the assessment mission proposed the following key benchmarks as preconditions for a more substantial drawdown of the force: (a) completion of credible disarmament, demobilization and reintegration of former combatants and dismantling of militias; (b) successful completion of the elections; (c) commencement of security sector reform, in particular confidence-building measures within a broader framework for democratic governance and oversight of the security sector, reunification of the army and establishment of functional and republican Ivorian armed forces and security services; and (d) restoration of State authority throughout the country.

48. In the meantime, UNOCI will continue to focus on helping the Ivorian parties to surmount the challenges and minimize the risks identified above, including by: (a) supporting the Ivorian parties and the Facilitator in their efforts to keep the peace process on track; (b) contributing to a secure environment for the elections, including by advancing the agendas for the effective disarmament of former combatants and dismantling of militias, inter alia, by monitoring the cantonment of former combatants and the storage of the weapons collected; (c) providing technical and logistical support for the electoral process and monitoring the progress made in the various processes leading to credible and transparent elections; (d) contributing, through my Special Representative’s certification role, to enhancing the credibility of the electoral process; and (e) supporting the economic recovery process pursued by the Government of Côte d’Ivoire.

In the SG’s next report (7 July 2009) he provided an update on progress against these benchmarks, this time as an Annex to the report (“Benchmarks and indicators of progress in key areas of the Ouagadougou Agreement and its supplementary agreements”), in list format. These benchmarks provide a systematic list of many provisions of the Ouagadougou agreement which remain outstanding, covering 1) DDR, 2) Elections, 3) Complete restoration of state authority and 4) Commencement of security sector reform.

The indicators themselves are fairly specific (e.g. “Profiling and cantonment of 5,000 Forces nouvelles elements at four sites in the north”), thus meeting the “clear and concrete” ideal. At the time, UNOCI was mandated to “support actively” implementation of outstanding OPA tasks (particularly elections and DDR), contribute to promotion/protection/monitoring of human rights and assist in reestablishment of a civilian policing presence throughout the country. Thus, the benchmarks are a welcome clarification of what the key aspects of the OPA are, as far as the UN presence is concerned.

However, in the Annex itself, the benchmarks do not distinguish who is responsible for their completion. While we know that profiling and dismantling of militias is “in progress,” it is not clear from the benchmarks who in fact is responsible for this. While the OPA doubtlessly lays out this responsibility more explicitly on the Ivorian side, a mandate to “support” the peace agreement leaves UNOCI in a gray area regarding their own responsibility, which these benchmarks do not resolve. The body of the report does provide
more information on UNOCI’s activities (what it is doing), but little on its responsibilities going forward (what it is expected to do). This remains the case through future SG reports.

UNMIT

The SG first produced benchmarks for UNMIT in his 4 February 2009 report on the mission, following the UNSC’s request in Resolution 1802 (2008) for him, “in consultation with the Government of Timor-Leste, to develop a medium-term strategy with appropriate benchmarks to measure and track progress.” The priority areas identified by the SG were:

1. Ensuring security and stability, now and in the future
2. Rule of law, justice and human rights
3. A culture of democratic governance and dialogue
4. Socio-economic development

The annex to his report lays out “objectives” and “benchmarks” under each area. Objectives tend to reflect more general goals, while benchmarks describe concrete indicators of progress against these goals. These benchmarks do not distinguish between core peacekeeper responsibilities and general targets. One of the Area 1 benchmarks provides an example: “Security sector institutions (including F-FDTL, national police, Segurança, Civil, customs, intelligence services and immigration) have adequate capacity (strategic, managerial and technical skills, as well as institutional values) to fulfill their roles and responsibilities, including human rights obligations.” Considering UNMIT’s mandate to assist the Government in this regard, further clarity is required concerning the tasks UNMIT is expected to accomplish itself. These benchmarks do not provide the needed clarity.

Furthermore, the benchmarks as presented are insufficiently specific. Another Area 1 benchmark is demonstrative: “The respective roles of the various institutions of the security sector are clearly defined.” This proscribes nothing about the eventual definition itself. Concerning the central importance of better defining these roles in Timor-Leste’s security sector reform process, this particular benchmark is not a useful tool for evaluating progress. Should the Government decide to remit key police responsibilities to the military, for example, evaluating progress against these benchmarks will not reflect this problem. Thus, these benchmarks are not even useful for a basic yes/no evaluation: by looking at the indicators provided, it is not clear whether the mission has completed its mandate.

---

22 To assist the Government in “working on a comprehensive review of future role and needs of the security sector, including the Ministry of Interior, PNTL, Ministry of Defence, and F-FDTL” (Res. 1802)