

Article

Evaluating national preventive mechanisms: a conceptual model

Richard Carver, Oxford Brookes University

Lisa Handley, Oxford Brookes University

Centre for Development and Emergency Practice

Oxford Brookes University

Oxford OX3 1BP

01865 482839 (P and f)

Dr Richard Carver is Senior Lecturer in Human Rights and Governance at the Centre for Development and Emergency Practice (CENDEP), Oxford Brookes University. Dr Lisa Handley is a Visiting Research Academic at CENDEP. Their joint publications include *Does Torture Prevention Work?* Liverpool: Liverpool University Press. 2016.

Summary

This article outlines a rigorous and systematic approach to evaluating both the performance and impact of national preventive mechanisms (NPMs) formed under the Optional Protocol to the United Nations Convention Against Torture. Many human rights practitioners remain sceptical about both the desirability and feasibility of evaluating human rights work. One obstacle has been that ‘indicators’ of human rights progress are formulated without evidence that they actually have a causal relationship to the intended outcome. By contrast, the tools used in this assessment model are derived from scientific research into what forms of torture prevention actually work, meaning that greater weight can be assigned to more effective activities (and vice versa).

The model is piloted in an assessment of the performance and impact of the Georgian NPM, which in ten years of work is shown to have had a significant impact in reducing the incidence of torture and other ill-treatment, particularly in police detention and prisons.

Keywords:

Evaluation, impact, national preventive mechanisms, Georgia, torture prevention

Funding

This work was supported by a grant from the Open Society Georgia Foundation.

Acknowledgments

The authors are grateful for the cooperation and support of the Public Defender's Office, especially Deputy Public Defender Giorgi Burjanadze, Nika Kvaratskhelia, head of the National Preventive Mechanism, and all other NPM staff, with particular thanks to Tamar Khokhobashvili and Akaki Kukhaleishvili.

Conflict of interest

The authors declare no conflict of interest with respect to this article.

1. Introduction

More than ten years ago, this journal published a special issue on measuring impact in human rights work.¹ Several contributors noted the growing pressure on human rights practitioners to submit their work to ‘monitoring and evaluation,’ while expressing great scepticism about both the desirability and feasibility of evaluating such work. A decade later, the pressure to measure the impact of human rights work has not lessened and there have been few positive developments to convince sceptics that it is possible and worthwhile to evaluate the results of their efforts. While it remains the case that activities such as human rights fact-finding and reporting are not readily susceptible to impact measurement, since outcomes usually occur so long after the original outputs, some types of human rights work can and should be evaluated. In this article, we outline a rigorous, systematic approach to assessing the work and impact of national preventive mechanisms (NPMs) established under the Optional Protocol to the United Nations Convention Against Torture (OPCAT) – an approach that we believe satisfies the demand for evidence-based monitoring and evaluation, while providing practitioners with a valuable tool for devising a more effective strategy to reduce torture and other ill-treatment in closed institutions.

Attempts to assess impact in the human rights field have fallen broadly into one of three categories. First, ‘human rights impact assessments’ have been directed primarily at entities, usually private companies, whose main activity is in another field but which has potential effects in the human rights field. This is analogous to (and overlapping with) the more established practice of environmental impact assessments. Secondly, human rights organizations of various types, non-governmental and statutory, find themselves increasingly subject to capacity assessments and evaluations. Finally, states are subject to scrutiny over their compliance in practice with their international human rights obligations.

¹ Vol I, Number 3, 2009.

It was the third of these, the aim of measuring states' compliance with human rights obligations, that led to the development of the human rights indicator framework (an approach later echoed in the extensive web of targets and indicators for the Sustainable Development Goals). Although the initial conceptualization of human rights indicators predated the 2009 special issue of this journal, the systematization of the indicator framework has been the most significant development in human rights impact assessment between that publication and the present. In the version developed by Office of the United Nations High Commissioner for Human Rights, human rights indicators are intended to measure three things: structural commitment to human rights, process in realizing rights, and outcomes.

Structural indicators are generally simple legal facts: whether a state is a party to, for example, the United Nations Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (UNCAT), and whether it has domesticated these treaty obligations into its municipal law (OHCHR 2012). Outcome indicators are also straightforward in principle. The number of reported cases of torture is a reasonable piece of data to collect, even if the number of *reports* of torture may raise as many questions as it answers (Clark and Sikkink 2013).

Process indicators are more problematic, at least in the context of torture prevention. They purport to measure the steps that a state has taken to convert legal (or structural) commitments into positive outcomes. In reality, they usually consist of a largely normative judgment or intuition about what a state should do to prevent torture, often untethered from any evidence about whether these mechanisms really work. Of course, this is in part because, until very recently, no rigorous, scientifically-based research was undertaken to determine which torture prevention mechanisms are actually effective and which are not. Our research, *Does Torture Prevention Work?*, a multi-country study that looked at torture prevention in 16 countries over a 30-year period (1985 – 2014), addressed this deficit (Carver & Handley 2016).

An example of misplaced reliance on a process indicator would be the OHCHR's first illustrative process indicator on the right not to be tortured: the proportion of complaints of torture received by ombudsman-type bodies and effectively responded to by government. Our research reveals that such complaints bodies have no discernible impact on the risk of torture, suggesting that this would be meaningless as an indicator of progress towards the eradication of torture. Equally, several of what we found to be the most effective predictors of reduced torture – notably safeguards upon initial arrest and detention – are not included in OHCHR's indicator framework (OHCHR 2012: 91).

We believe that any assessment of the long-term impact of a torture prevention programme should focus on the adoption of those mechanisms that have been shown to be the most effective at reducing torture, rather than those simply intuited to be the most important. Our evaluation model not only identifies what preventive measures have been adopted but weighs those with a demonstrable effect on torture reduction more heavily than those that do not in the final assessment of the torture prevention programme. For example, a preventive strategy that leans heavily upon complaints mechanisms will score poorly in our evaluation scheme, while one that emphasises protections against incommunicado detention – found in our research to be the most important – will score highly.

The research underlying *Does Torture Prevention Work?* addressed the issue of how to stop torture and other purposive ill-treatment, not all forms of ill-treatment. This is narrower than the mandate of NPMs, which seek to prevent all forms of ill-treatment of persons deprived of their liberty. However, the plain text of the OPCAT (Article 17), along with the UN Convention itself, make it clear that the primary purpose of 'national preventive mechanisms for the prevention of torture' is, unsurprisingly, to prevent torture. In the vast majority of States Parties to the OPCAT torture remains an endemic problem. In the remainder, it is still the duty of the NPM to prevent its occurrence. Hence, the impact framework described in this article is directed towards the role of NPMs in preventing torture and other purposive (or deliberate) ill-treatment, and hence has more relevance to a country like Georgia, where the risk of

torture remains a constant threat, rather than a country such as Norway, also in our original study, where it is not.

Discussing how to evaluate national human rights institutions (NHRIs), Julie Mertus identifies three broad approaches: structural, mandate and impact (Mertus 2012). Structural approaches focus on compliance with the Paris Principles relating to the status of national institutions, that is with a legal-normative framework. Mandate approaches assess the degree of compliance with NHRIs' own remit, which may differ from the Paris Principles. Mertus's third and preferred approach is to evaluate the impact that NHRIs have on the long-term promotion and protection of human rights. In practice, NHRI evaluations have tended to focus on the first two, for the simple reason that determining external impact, whether short- or long-term, is such a fraught exercise. How is it possible to distinguish the impact of NHRIs from other unconnected factors that may affect human rights? How can we make such an assessment when the range of relevant human rights is so broad?

Evaluating NPMs – many of them, of course, housed within NHRIs – is slightly more straightforward since only one set of rights is involved. Now, based on our research, which clearly distinguishes the impact of different variables on torture prevention, it is also possible to identify the specific contribution of NPMs.

In this assessment, structural and mandate elements remain important. The ability of the NPM to conduct its work (for example, whether it has the resources and independence needed), the actual work carried out by the NPM (for example, whether monitoring visits are frequent and unannounced, whether recommendations for improvements are offered), and the short-term outcomes of this work (for example, whether the recommendations are adopted), play a role in how effective the NPM is at reducing torture and ill-treatment in the long term. The evaluation model described in this article considers all four of these components; components referred to in traditional programme evaluation

language as the inputs, outputs, and short-term and long-term outcomes of the work of the NPM. Three of the components, inputs, outputs and long-term outcomes, are evaluated at least in part using an assessment tool derived from *Does Torture Prevention Work?* – a tool we call ATEMPT. Another segment of inputs is measured based on a tool we developed for another project. These two assessment tools are described in Section 3; the NPM evaluation model itself is illustrated and briefly described in Section 2.

What we propose is a broad framework for assessing NPM effectiveness that looks at the formal and structural characteristics of the institution, how it works, and its impact on the incidence of torture and ill-treatment. Different assessment tools could be used to evaluate different elements in this process. For example, both the Association for the Prevention of Torture (Buckland & Csergő 2019) and Ludwig Boltzmann Institute for Human Rights (Birk and others 2015) have made very useful suggestions on how to track and measure the impact of NPM recommendations. Such approaches could easily be incorporated into our proposed model. By the same token, more detailed information could be collected on the inputs and outputs of NPMs. The point of the model we offer here is twofold. First, an evaluation should include an examination of not simply the structure and mandate of the NPM (or the inputs and outputs in our framework), but also the impact, that is, the short-term and long-term outcomes. Second, any evaluation must rest on assessing factors that truly are related to a decrease in torture, not on factors that we intuit are related to torture prevention.

One of the countries included in *Does Torture Prevention Work?* was Georgia. Georgia is often regarded as having made significant strides in torture prevention in recent years and, in 2019, we returned to Georgia (our earlier research focused on 1985-2014) to see if we could substantiate this claim and determine what role the NPM, established ten years earlier, played in this progress. It is in this context that we devised our overall NPM evaluation model. And in Section 4 we discuss our application of this model to the NPM to ascertain its impact in reducing torture and ill-treatment in Georgia.

2. Conceptual Model for Evaluating NPMs

Our framework for assessing the effectiveness of an NPM is a modification and expansion of the traditional programme evaluation logic model, incorporating inputs, outputs, and outcomes. Our model identifies and evaluates the inputs and outputs of the NPM, as well as the short-term and long-term impact of the NPM on torture prevention.

[Figure 1 about here](#)

The first component of the evaluation model displayed in Figure 1 is an assessment of the independence, resources and general capacity of the NPM to do its work – that is, does it have the ‘inputs’ required to carry out its work effectively? The second component we consider is ‘outputs,’ that is, the actual work of the NPM: How frequently does it visit closed institutions? Are these visits unannounced? Are interviews of detainees conducted and are these private? Are reports of the NPM’s findings published? These ‘inputs’ and ‘outputs’ are the portion of the conceptual model referred to as ‘process.’

The second set of components in the conceptual model in Figure 1 is ‘outcomes’ and this includes both the short-term outcome and the long-term impact of the work of the NPM. In the short term, the effects of the NPM are best assessed by reviewing the recommendations it has put forward and whether these recommendations have been implemented. Of course, in the long term, the goal of these recommendations, and the work of the NPM in general, is to reduce torture and ill-treatment. These relationships are represented by the arrows that point from adoption of recommendations and from monitoring in general (the general operations of the NPM) to the long-term goal of reduction in the risk of torture.

The impact of monitoring bodies on reducing torture and ill-treatment may be indirect as well as direct. Recommendations that seek to improve detention and prosecution laws and practices are not only relevant as potential short-term outcomes of the NPM, but also improve detention and prosecution practices, both of which play independent and essential roles in preventing torture and ill-treatment. Hence our model includes arrows from the adoption of recommendations to detention and prosecution, indicating the possible effect of NPM recommendations on detention and prosecution law and practice. Sitting outside this model, and potentially influencing all parts of it, is the specific country context. Political, social, and legal factors will affect the incidence of torture, but also the functional independence of the NPM, available resources, and the willingness of authorities to comply with its recommendations. These factors are largely beyond the control of the NPM itself. To give an example from our earlier research: Turkey significantly reduced the incidence of torture in the 2000s. The political motivation behind this was the country's application to join the European Union, a factor that was country-specific. The mechanism by which it was achieved was a radical overhaul of detention law and practices, so that suspects were far better protected on arrest. Later, again for political reasons, the Turkish government reversed this process in order to facilitate the torture of political opponents. Similarly, improvements in the protection of inmates from torture in Georgia have been driven by two important political events: the Rose Revolution of 2003 and the 2012 election of the Georgian Dream party, following a national scandal about torture in prisons. Each of these political transitions set in motion legal and procedural reforms that have led to a reduction in torture.

3. Sources for Measuring Components of the Evaluation Model

Our model for evaluating the effectiveness of NPMs relies in part on two assessment tools developed as part of previous research projects. One tool (RIPE) originated in our work for the United Nations

Development Programme to develop a tool for ascertaining compliance with the Paris Principles. We use a modified version of this tool to evaluate NPM inputs. The second assessment tool (ATEMPT) was derived from our multi-country study on the effectiveness of torture prevention mechanisms, published as *Does Torture Prevention Work* (Carver and Handley 2016). A portion of this tool assesses monitoring law (inputs), monitoring practices (outputs) and, in its entirety, ATEMPT evaluates a country's overall anti-torture strategy.

3.1 Inputs: RIPE

We refer to our assessment tool for evaluating the independence and resources of NPMs as Resources, Independence, and Pluralism Evaluation, or RIPE. It is based upon an effectiveness framework that we developed on behalf of the United Nations Development Programme for evaluating national human rights institutions (Carver 2014). The original framework looked at the relationship between, on the one hand, various structural and legal attributes usually regarded as being important for NHRIs and mainly deriving from the Paris Principles, and, on the other hand, various important outputs. This is primarily a tool for assessing 'first-order effectiveness' – that is to say, the activities and outputs of the institution, rather than the overall impact of the NHRI on society. We had to adapt this tool since it addresses itself to the attributes and activities of the NHRI. A national preventive mechanism is not, in itself, a national human rights institution, even though it may be housed within one. Therefore, it does not have the full range of attributes and functions that an NHRI should have. However, the OPCAT does specify that states should give 'due consideration' to the Paris Principles in creating national preventive mechanisms (Art. 18(4)). From the context, this primarily refers to the provisions of the Paris Principles relating to independence, pluralism and the need for adequate resources (Art. 18(1-3)).

RIPE entails, in effect, a capacity assessment of the NPM, ranging from the legal framework within which it was created, to the resources available to carry out its mandate. The first set of RIPE criteria

specifically relate to the independence of the NPM and include such factors as the statutory basis for the NPM, the appointment and removal processes, the criteria for membership, the term of office, adequate remuneration and protections against instruction from government. The second set of RIPE criteria relates to the resources available to the NPM. Does it have adequate staff? Do they receive adequate training? Does the NPM have control over its own budget? Is it adequately funded? The final section of RIPE addresses accessibility and partnerships with civil society. Is there formal consultation with civil society? Are there joint activities with civil society organizations? Does the NPM have local offices? Each of these issues is answered using a codebook consisting of multiple choice questions and an overall score assigned.

We also use the monitoring law component of ATEMPT (see below) to measure the legal inputs to NPMs, including its powers to conduct unannounced visits and to conduct confidential interviews with persons deprived of their liberty.

3.2 Outputs and Long-term Outcome: ATEMPT

A few years ago, the Association for the Prevention of Torture approached us with a question: Do the various mechanisms put forward as preventive actually reduce the risk of torture? After four years of work by more than 20 researchers in 16 countries, we determined that most of the requirements of international law with regard to torture did have a positive effect in the countries that we studied – anti-torture campaigners and international lawyers had not, by and large, been wasting their time supporting these preventive measures (Carver & Handley 2016). However, the most interesting aspect of our findings was that the order of priority assigned by many anti-torture campaigners may not have been correct. We grouped anti-torture measures into four broad categories. The first was investigation and prosecution of torturers (as required by the UNCAT). The second was the establishment of independent monitoring bodies at a national level, and granting access to international and regional monitoring

bodies (as required by the OPCAT).² The third was the establishment of non-judicial complaints mechanisms (as recommended by the European Committee for the Prevention of Torture and strongly promoted by various UN bodies).³ The fourth was the various procedural guarantees that should be in place when a person is taken into custody (as recommended, but somewhat weakly, by the UN Body of Principles).⁴ It turned out that detention safeguards, the least strongly entrenched in international law, were by some distance the most important in protecting against torture. Prosecution of torturers was also important and the establishment of independent monitoring mechanisms had a discernible positive impact. Independent non-judicial complaints mechanisms did not have a detectable effect in preventing torture in the countries that we studied, although they may provide some sort of remedy at an individual level, for example in claims for monetary compensation or other forms of redress.

Our other principal finding also has significant implications. For each of our four clusters of preventive mechanisms – detention, prosecution, monitoring and complaints – we looked both at how far each measure existed in law, but also whether it was fully implemented in practice. We found very clearly that there was a wide gap between law and practice, especially with regard to detention safeguards and prosecutions – the two most effective sets of preventive measures. Indeed, we saw that adopting any of these preventive measures in law had no effect in reducing the incidence of torture; this only happens when they are implemented in practice. Of course, enacting prevention into law is usually (though not

² The Optional Protocol requires states parties to open the doors of their prisons and detention centres to an international monitoring body, the Sub-Committee on Prevention of Torture, as well as creating *national* preventive mechanisms as independent monitoring institutions. Similar provisions exist at the regional level, notably the European Convention for the Prevention of Torture

³ The international promotion of national human rights institutions has led to the proliferation of non-judicial complaints mechanisms, seen by many as an important part of torture prevention.

⁴ Various non-binding international instruments, notably the United Nations Body of Principles for the Protection of All Persons under Any Form of Detention or Imprisonment and the Nelson Mandela Rules (the Standard Minimum Rules for the Treatment of Prisoners) establish important standards and procedural guarantees, many of them given stronger expression in the jurisprudence of the regional human rights courts, especially in Europe.

invariably) a necessary precursor to implementing prevention in practice. Table 1 provides an example of some of the preventive mechanisms we considered by cluster.

[Table 1 about here](#)

Within our four groups of preventive mechanisms, it was also possible to identify which specific safeguards were most important. Among the most important were that people only be held in official, recorded detention; that members of their family or another person be promptly notified of their detention; that they be informed and exercise their right to a lawyer promptly after arrest; and that they receive an independent medical examination in custody. More broadly, criminal justice systems that rely more on alternatives to confession evidence are likely to see reduced torture.

With regard to the prosecution of torturers, the most important step is that complaints of torture are actually lodged with the appropriate prosecutorial authority. The inference here is that very often victims of torture have no confidence in the system either to investigate properly or to protect them against further abuse. Allegations of torture need to be fully and impartially investigated; conviction rates need to be comparable to those of other serious crimes; and there must be no amnesty or pardons for torturers.

The most important element of effective monitoring is the functional independence of the monitoring body – when monitors and those detainees that cooperate with them are free from threats or sanctions, then there is a higher correlation with reduced torture. It is also important that monitoring bodies conduct confidential interviews with persons deprived of their liberty. (All NPMs should have the power to do this, under Article 20[d] of the OPCAT.)

Monitoring bodies may also have an important indirect impact in reducing torture through the recommendations that they make to the authorities. If their recommendations correspond to the most important steps needed to prevent torture – and if the recommendations are then implemented – they

can be a valuable part of the preventive architecture. In other words, one would expect that NPM recommendations that had a serious impact would be largely strategic. They would not merely react to particular bad conditions that were observed, but propose laws, policies, and systems that would reduce the risk of torture and ill-treatment.

To research *Does Torture Prevention Work?* we used a combination of methods. We conducted 16 qualitative case studies of countries over the period 1985 to 2014. In each country we looked at the incidence of torture and the various legal and policy steps taken to prevent it. This was placed in the context of the political and social issues in each country to explain the drivers behind torture and ill-treatment, as well as the motivations or obstacles to preventive measures. However, our overall conclusions about the efficacy of prevention were drawn from a statistical analysis that measured the impact of various preventive mechanisms on the incidence of torture. For each of our 16 countries we looked separately at each of the 30 years (1985-2014) and determined if the preventive mechanism was present (or partially present or absent). We did this for more than 60 preventive measures grouped into the four clusters described above. We also did this for the dependent variable: the incidence of torture measured using a new scale called the Carver Handley Torture Score – CHATS.

The statistical analysis allowed us to determine which preventive mechanisms were most effective at reducing torture. We used this information to develop an assessment tool to evaluate a country's anti-torture strategy. This is called the Assessment Tool for Evaluating Mechanisms for Preventing Torture (ATEMPT). Most assessment tools are simply a list of questions. An example of such a checklist is the 'assessment matrix' devised by the UN Sub-Committee on Prevention of Torture for evaluating national preventive mechanisms (SPT nd). For our list of questions, however, we have the advantage that we have already tested and drawn conclusions as to which answers matter more in reducing torture. And we use this information to weigh those mechanisms and clusters of mechanisms that have proved the most effective in reducing torture more heavily than those that are less effective.

The ATEMPT questionnaire mirrors our codebook for the multi-country study. It consists of 63 questions corresponding to 63 preventive measures. The questions are divided into four categories – detention, prosecution, complaints, and monitoring – and each of these clusters is then further subdivided into sections on law and practice. The monitoring portion is also divided between police detention and prisons. The questions are multiple choice, usually with three options (the preventive mechanism is wholly present, partly present or wholly absent).

Each question is scored depending on the answer and then weighted, depending on how important that particular mechanism has proven to be in preventing torture. The higher the correlation between the given mechanism and its role in reducing torture, the greater the weight assigned.

The weighted individual scores within each category are then tallied to produce a cluster score. In order to compare scores across categories/clusters, the scores are expressed as a percentage of the total points possible for each cluster. For example, the highest score possible for the detention cluster, when weighting is taken into account, is 78. If the detention items tally to 43 for a given country, the score for that cluster would be 55.13% (since 43 is 55.13% of 78).

To produce an overall score, a second weight is then applied to each cluster score, depending on the overall importance of the category in preventing torture. Again, the higher the correlation between the cluster and the reduction in torture, the greater the weight assigned. These four weighted cluster scores are then summed for the total score. Countries will score higher when they put their efforts into achieving those measure that have been found to have the strongest impact on torture prevention.

One component of assessing the overall state of a country's torture prevention system is to assign scores to the mechanism(s) designated to monitor closed institutions. ATEMPT questions relating to inputs include: Is there a domestic monitoring mechanism established by law? Does the mechanism have the power to make unannounced visits to places of detention? Does the monitoring mechanism

have the power to conduct interviews with inmates? Is the monitoring mechanism required to issue reports on its activities? Do monitors have legal immunity from sanction for their work? We use the monitoring practice component of ATEMPT to measure the output of the NPM, including: Does the domestic monitoring mechanism conduct regular and frequent visits? Does the domestic monitoring mechanism conduct unannounced visits? Does the monitoring mechanism conduct interviews with detainees? Does the domestic monitoring mechanism publish its findings? Have domestic monitors been sanctioned for their monitoring-related activities?

3.3 Short-term Outcomes: Tracking Recommendations

Systematically tracking NPM recommendations and whether they have been implemented is crucial for determining the short-term impact of an NPM. Of course, the fact that a change recommended by the NPM has been implemented may not be attributable solely to the NPM, but it unquestionably plays a role in this process. Equally, implementation of recommendations may, in part be attributable to the quality of the recommendation. This is customarily assessed by SMART criteria; is the recommendation Specific, Measurable, Assignable, Realistic, and Time-related (Doran 1981)? Later versions are usually some variation on Specific, Measurable, Achievable, Relevant, and Time-bound (Yemm 2013). Yet NPM recommendations do not necessarily comply with all these criteria. For example, some recommendations of the Georgian NPM did not meet the standard of Realism/Achievability, since they could not be immediately met within the available budget. Yet such recommendations – including the closure and rebuilding of prisons – have been fulfilled eventually in some instances. More often, we suspect, non-compliance with a recommendation will be primarily a result of political factors or other causes related to the country context, rather than the quality of the outputs of the NPM.

According to the Association for the Prevention of Torture (APT), NPMs use a variety of tools to track their recommendations: ‘Many NPMs use an excel spreadsheet. Others have had a database custom

built for them. Some NPMs are part of a larger institution and therefore use the broader institution's existing data systems.' (Buckland and Csergő 2019: 14) The key is to ensure they are 'not overly resource-intensive to use' and are therefore maintained over time (Ibid: 14). As important as identifying the recommendation, in our view at least, is devising a method for 'categorising the recommendation into implemented, partially implemented, not implemented.' (Ibid: 16)

Because our evaluation model was devised in conjunction with assessing the impact of the Georgian NPM, which had only recently begun to systematically track recommendations, this component of our evaluation model is less developed than the other three. We hope to address this deficiency in future research.

4. Application of Evaluation Model to Georgian NPM

Our evaluation model, incorporating the two assessment tools described above, was developed in response to a request by the Georgian NPM to assess its effectiveness in preventing torture and ill-treatment on the occasion of its tenth anniversary in 2019. Since Georgia was one of the 16 countries included in *Does Torture Prevention Work?*, we had already collected much of the relevant information from 2009 through 2014.⁵ In-country research in 2019 allowed us to update and expand our study using the tools we describe in this article.

Georgia was one of the first countries to ratify the OPCAT and hence one of the first to be faced with the challenge of how to constitute a national preventive mechanism. Georgia made the same decision as the overwhelming majority of parties to the OPCAT and designated its (internationally accredited) national

⁵ The researchers who wrote the chapter on Georgia in *Does Torture Prevention Work?* were Bakar Jikia and Moris Shalikashvili. We are very grateful to them for work that provided the foundation for this further study.

human rights institution, the Public Defender's Office (PDO), as the NPM. This had various practical advantages: the PDO already had the required legal powers (although the Organic Law on the Public Defender was amended to make its new function explicit) and had been actively engaged in monitoring closed institutions for a number of years. Hence, if any Georgian institution had the necessary expertise, it was the PDO.

The revised Law on the Public Defender created a body called the Special Preventive Group, which encompassed both the staff of the NPM itself and a broader range of experts to be recruited and empowered to act as members of the NPM as required. In addition, an advisory council was created, comprising members of civil society and other experts, which was intended to give broad strategic guidance to the NPM. The decision to adopt this model was made in 2008 and the NPM became operational in 2009.

Initially, responsibility for the functions of the NPM rested with the department already dealing with criminal justice and penitentiary matters, since this was where the expertise for monitoring visits rested. The number of personnel in this department was increased, but resources were still inevitably limited, since the department had responsibility not only for preventive monitoring visits, but also the processing and consideration of large numbers of individual complaints from persons in police or prison custody. The recruitment of the Special Preventive Group (SPG) did, however, constitute a significant expansion of the PDO's capacity, both in simple numbers available to conduct monitoring visits, but also in expertise. This was particularly apparent in relation to disciplines such as medicine, not much in evidence within the PDO, but available for hire through the SPG. This was important in relation to a key part of the NPM's work – monitoring of psychiatric hospitals – but also in relation to evaluating prison healthcare, which has been a continuing priority over ten years, and ensuring that the skills exist in the detention isolators and prisons to conduct forensic medical examinations to document allegations of torture or other ill-treatment.

For approximately the first half of the NPM's life, preventive monitoring and complaints handling remained functions of the same department. This is visible in the very high numbers of visits recorded in the period 2009-13, which dropped dramatically from 2014 onwards. The reason for this was not a change in the number of preventive visits, but the fact that the NPM no longer conducted visits to address individual cases. Under the old system, a half-hour meeting with a complainant, which would usually not even entail the NPM staff member entering the main part of the closed institution, would nevertheless count as a visit.

In our interviews with outside actors, whether governmental or non-governmental, no one made any distinction between the NPM and the PDO. Indeed, all our questions were framed using the term 'NPM' but were almost invariably answered with reference to the 'PDO.' This is not to suggest that outsiders do not understand the distinct nature of the NPM function, but it is clear that the NPM is not seen as a different entity from the PDO. While there is potential for confusion, this is not a bad thing. When the NPM was assigned to the PDO, the office and its incumbents enjoyed considerable popular trust and a reputation for being willing to challenge the authorities (NDI 2019). This sort of public legitimacy can be immensely important to the credibility and impact of the NPM.

There may be overlaps between the work of the NPM and that of other arms of the PDO. For example, although the NPM no longer handles complaints, this does not mean that the PDO does not do monitoring. A positive example of this was described by a staff member of the PDO regional office in Batumi. The office receives very few complaints relating to treatment in police custody, in contrast to inmates in the penitentiary system. The staff member conducts monitoring visits to police stations and temporary detention isolators, opening a complaint file if he identifies a breach of a suspect's rights. This is part of a more general policy of conducting proactive visits to identify human rights violations in closed institutions, not only police facilities, that have not been the subject of a complaint. This is clearly a different process from that conducted by the NPM, but it supplements the NPM's work in an

important way. The NPM is entirely based within the national headquarters in Tbilisi. Between visits, the staff of the regional offices are the sole presence of the PDO. Given that the frequency of NPM visits to local police stations is going to be very low (though much higher for temporary detention isolators), this is an effective way of maintaining a monitoring presence at an important site of risk to persons deprived of their liberty. (Batumi police, in our interview with them, made no distinction between the NPM and the PDO.)

4.1 Inputs: RIPE and ATEMPT

We used a combination of two tools to evaluate the inputs into the Georgian NPM's work. RIPE measures compliance with the provisions of the Paris Principles relating to independence, pluralism and the need for adequate resources. The ATEMPT monitoring law component measures the legal framework governing the NPM's work.

4.1.1 Independence

The independence criteria address the statutory basis for the NPM, the selection and appointment process, criteria for selection, security of tenure, conflicts of interest and protection against coercion, and immunities enjoyed by both the head of the institution, who in law is the Public Defender, and its members, who are the experts of the Special Preventive Group. The PDO scores almost perfectly on these criteria, with a constitutional and statutory foundation, strong guarantees against interference, and also immunities for acts carried out in the course of their work. The Georgian NPM scores slightly less than perfectly because the Organic Law on the Public Defender does not lay out explicit human rights criteria for eligibility. The score is 17/18 or 94.4%.

4.1.2 Available Resources

Here we looked at staffing: was it adequate? Was the NPM free to recruit its own staff? And did they receive adequate training? We also considered budget and financial resources: did the NPM control its own budget? Was it adequate? And was the NPM free to raise funds from elsewhere? The Georgian NPM does not have many staff, but the fact that it draws on the external professionals of the SPG means that it has adequate human resources, whom it recruits independently. While the PDO has a separate budget and is adequately funded, the budget for the NPM is not guaranteed by law. In practice, while the state budget covers most NPM salaries, it does not cover operational costs, which are met by private donors. When we did our research, a financial crisis was looming, with an important long-term donor likely to withdraw. The total score on the resources criteria is 9/12 (75%), primarily because of budgetary issues and some weaknesses in relation to training.

4.1.3 Accessibility and Partnerships with Civil Society

This section of RIPE addresses the accessibility of the NPM, for example through local offices, and its partnerships with civil society organizations. Although the NPM department itself does not have staff in the regions, the PDO has multiple local offices with staff who collaborate closely with the NPM and conduct some monitoring of their own. The SPG, which provides the expertise for monitoring visits, is a formalized civil society partnership involving academia, the medical profession, and others. The NPM also conducts joint projects with NGOs. The Georgian NPM scores 100% (6/6) on these criteria.

The overall RIPE score is 32/36 or 88.9% with a perfect score for civil society relations and the weakest score (75%) for resources.

4.1.4. Monitoring law (ATEMPT)

ATEMPT provides a score for monitoring law (as well as for monitoring practice, which is an *output* of the NPM). This includes the powers available to the NPM: to conduct unannounced visits, to conduct confidential interviews with inmates, and to issue reports and recommendations on its findings. The Georgian NPM has all the necessary legal powers, except an explicit protection against sanction for those who communicate with the institution. Its score is therefore 92%.

4.2 Outputs: ATEMPT Monitoring Practice Score

The principal outputs of an NPM are its visits to closed institutions and the reports, recommendations, and proposals that derive from its monitoring and other research. The ATEMPT monitoring practice score measures the most important of these: regular and frequent visits, often unannounced, confidential interviews with inmates, reports and recommendations, and practical immunity from sanction. The Georgian NPM scored the maximum on this. It has an extensive visiting programme, with all visits unannounced. In the period prior to the creation of the NPM, there were serious threats of reprisal against some PDO employees as a consequence of their monitoring work. This has not happened in the past ten years and NPM staff and SPG members have been able to work unimpeded.

As noted above, recommendations are conventionally supposed to comply with SMART criteria. Yet recommendations in the human rights field often have a more programmatic or aspirational character. For example, when the Georgian NPM recommended the closure of a prison or the drastic reduction in the number of prisoners, these may not have been 'realistic' or 'attainable' according to the SMART approach but they ultimately set a new agenda and led to change that benefited prisoners. Likewise, in its early years NPM reports consistently documented the incidence of torture in prisons at a time when the issue was not publicly acknowledged and recommendations were ignored. In 2012, a public scandal involving the release of a video of torture forced a new government to take action; many NPM recommendations were adopted and its monitoring reports vindicated. The point is that without these

changed political circumstances, the recommendations might never have been adopted – but they were still the correct recommendations. No monitoring body is infallible, but there is a reasonable expectation that monitors will accurately identify the most serious problems. On this the PDO, both before and since its designation as NPM, has a strong record. The tendency of some NPMs to focus on conditions of confinement to the exclusion of monitoring for torture may lead to recommendations that are more ‘realistic’ – and hence more likely to be implemented – but which are actually poor outputs because they have failed to identify the most serious issue.

The Georgian NPM scored 100 per cent on the ATEMPT monitoring score.

4.3 Short-term Outcomes: The Impact of NPM Recommendations in Georgia

To determine the impact of the Georgian NPM on reducing torture and ill-treatment – the right-hand side of our conceptual model – we consider both short-term outcomes and the long-term impact. We believe the impact of the NPM in the short-term is best gauged by looking at what NPM recommendations have been adopted. As a consequence, in this section we briefly discuss the recommendations issued by the NPM.⁶ However, because the NPM has only recently begun to systematically track whether recommendations have been adopted, our assessment of the short-term impact of the NPM is less structured than might otherwise be the case.

As noted above, improvements in the treatment and conditions of persons in closed institutions are always going to be determined by a variety of factors. The mere fact that an improvement was made and the change was previously included in a list of NPM recommendations is not itself proof that it was the NPM’s advocacy that secured the change. Nevertheless, it is undeniable that the NPM in Georgia is engaged in a process of formulating policy proposals for prisons, police custody, temporary detention

⁶ A much more in-depth discussion of the recommendations issued by the NPM in Georgia and their impact can be found in Carver & Handley (2019).

isolators, and psychiatric hospitals that goes far beyond what is offered by any other non-executive agency. We identified an increasingly close engagement and dialogue with the relevant government authorities, which was the result of a conscious strategy on the NPM's part. The recommendations issued by the Public Defender as NPM are included in the annual report to Parliament and considered there by the Committee on Human Rights. Recently, the Committee has adopted the practice of endorsing many of these recommendations, a practice welcomed by the NPM and seen as making them more likely to be implemented. We were sceptical, since this practice seemed to create two tiers of recommendations, with the likely result that those not endorsed by Parliament would be ignored.

Non-compliance with recommendations does not necessarily mean rejection. Many NPM recommendations have entailed substantial spending that would not have been possible in the following year. However, many such recommendations, including the closure of penitentiaries, have been implemented in the long run. Most importantly, the drastic overcrowding of prisons has been reversed by a systemic policy of early release combined with greater use of non-custodial sentences, halving the prison population and contributing to a substantial improvement in conditions.

As noted above, the NPM documented torture in the prisons before it became a matter of public controversy and the government took action. One clear obstacle that it identified, still largely unresolved, is effective impunity for acts of torture. The government has finally made an institutional reform that may lead to improvements – the creation of a special investigative office, in compliance with longstanding PDO recommendations – but its efficacy remains to be seen.

In many other areas, there have been significant improvements in prisons that can be traced back to NPM recommendations. These include the introduction of new procedures based upon the Istanbul Protocol to document cases of ill-treatment, with medical doctors trained in their use. Prisons are now required to retain closed circuit television footage for at least one month (longer if it is required as

evidence), whereas when the NPM was established CCTV evidence would invariably be wiped. The NPM has succeeded in securing reforms of disciplinary and security procedures that constituted ill-treatment, including ending confinement to 'de-escalation rooms' without time limit. Material conditions have improved substantially, mainly because of the reduction in the prison population. Healthcare in prisons remains a major problem, although the NPM has succeeded in pressing for a more preventive approach, including screening for various diseases and detoxification programmes. Mortality rates within the prison system have dropped dramatically.

Before the creation of the NPM, a system of temporary detention isolators was created to house newly arrested suspects and address the problem of torture in police stations. The new approach was dramatically successful; the NPM continues to monitor both the isolators and police stations, where excessive use of force continues to be a problem. At the NPM's initiative, proper detention logs are now kept and the use of CCTVs in police stations is more systematic.

4.4. Long-term outcome: the role of the NPM in reducing torture in Georgia

Because Georgia was one of the countries included in *Does Torture Prevention Work?*, we can use ATEMPT to compare the strength of the preventive mechanisms in place in Georgia in 2008 (before the establishment of the NPM) and, based upon our new research, in 2018, as well as comparing the incidence of torture both before the establishment of the NPM and in 2018-19. On the latter point, it is apparent, based upon a wide variety of sources that, using our own Carver Handley Torture Score (CHATS), torture and ill-treatment in Georgia have declined from a high of 5 to a score of 2.⁷ The incidence of torture in police custody dramatically declined in the mid-2000s, before the establishment of the NPM, although the PDO played an important part in this. Torture in prisons, as well as very poor

⁷ The CHATS measures the incidence of torture on a scale combining frequency, severity and geographic dispersal. It is calculated on the basis of all available primary and secondary sources. It is fully explained in Carver and Handley (2016), pp. 39-42.

conditions and other ill-treatment, was at a very serious level in the late 2000s, but is now much less frequent, widespread and severe.

Figure 2 reports the detention, prosecution, complaints, and monitoring scores for Georgia for 2008 and 2018. The higher the score, the more preventive mechanisms in that category in place, and the more effective those mechanisms are for reducing torture. The highest score possible is 100% – meaning that all of the preventive mechanisms identified by ATEMPT are in place. Lower scores mean that there are fewer mechanisms in place and/or the mechanisms that are in place have been found to be less effective in reducing torture than the ones that have not been put in place.

What is clear from comparing the scores in Figure 2 is that (1) the monitoring and complaints scores are much higher than the prosecution and detention scores and (2) there has been no change in the complaints or prosecution scores between 2008 and 2018. In this instance at least, ATEMPT allows us to do something that is always a great challenge when it comes to assessing the impact of any particular human rights body, such as a national human rights institution. Because the scores in two of the four categories (prosecution and complaints) remained constant over the ten-year period, any reduction in the incidence of torture is most likely to be a consequence of improvements in detention and/or monitoring as opposed to prosecution or complaints.

[Figure 2 about here](#)

As indicated in Figure 2, the scores for monitoring and complaints in 2008 were both high (82.74% and 82.14% respectively), reflecting both the strong legal basis of the PDO and its good practice in both its complaints and monitoring functions. However, as noted above, the complaints function is found overall to be of relatively little significance as a measure to prevent torture and ill-treatment. The scores for detention and prosecution, both of which are more important for prevention, were lower than the monitoring and complaints scores in 2008 (55.13% and 55.36%).

The detention score reflects the introduction of some strong procedural protections in law prior to 2008, but also some continuing gaps, for example in the use of video monitoring and electronic recording of police interviews. It also reflects some weaknesses in practice: while families were almost always promptly informed of a person's detention and the detainee was brought before a judge within the legal time limit from the formal time of detention, there continued to be some use of unofficial detention – for example, in manipulating records of the time of arrest – and detainees were not always informed of their right to a lawyer.

On prosecution, the score also reflects a gap between law and practice, with torture and deliberate ill-treatment criminalized, but seldom properly and effectively investigated or prosecuted. Prosecutions of alleged torturers were few, as were conviction rates and officials convicted of such crimes could expect to face lower penalties than equivalent civilian criminals.

The maintenance of the high complaints score reflects the fact that in both law and practice Georgia in 2008 already had a highly developed non-judicial complaints mechanism in the PDO. The only weaknesses identified in our scoring related to the legal powers of the PDO to compel the production of evidence and witnesses and its powers to recommend redress. None of this changed between 2008 and 2018 – the amendments to the Organic Law on the Public Defender in 2009 and 2010 did not affect these points.

The low prosecution score reflects almost entirely the actual practice (as opposed to the law), with a general lack of investigations of those alleged to have committed torture or ill-treatment. In the rare event that prosecutions are initiated, the rate of conviction is far below that for equivalent crimes committed by ordinary criminals and, even on conviction, sentences are much lower. Despite a number of prosecutions relating to well-publicized acts of torture in prison that were exposed in 2012, the general pattern has not changed between 2008 and 2018. This is an issue that has been a constant focus

of recommendations from the PDO, dating from before the formation of the NPM. The creation of a new special investigator's office with a mandate to investigate (but not prosecute) allegations of torture and ill-treatment might make a difference in this. This proposal has been adopted by government but, since it has not yet been implemented, we cannot determine its effect.

The detention index, we know from our multi-country study, is the most important in terms of impact on the incidence of torture. The improvement from 55.13% in 2008 to 60.26% in 2018 is explained by two changes. One is the introduction of mandatory medical examinations for detainees on transfer from police custody to temporary detention isolators. The other, which is directly attributable to the recommendations of the NPM, is the increased use of closed circuit television to monitor police stations and temporary detention, as well as a significant improvement in the retention of recordings (from 24 hours to approximately one month) and better practice in making the recordings available in the event of an investigation.

The monitoring index directly measures the work of the NPM, since the PDO is the official monitoring body for both police and temporary detention on the one hand, and penitentiaries on the other. (We measure the two separately because in some countries different bodies are assigned to different types of deprivation of liberty.) The monitoring score for 2018 is nearly perfect (97.44% out of 100%), reflecting both a strong legal framework and good practice. From the perspective of our scoring, the legal framework did not change between 2008 and 2018, although the 2009 amendments to the Organic Law on the Public Defender provided the legal basis for the NPM. That law established the Public Defender as a monitoring body appointed by Parliament with strong guarantees of independence, both in the appointment process and in the legal protections against interference in its work. Although the 2009 amendments specifically designated the PDO as the NPM, there were already strong guarantees of independence through the appointment process and the prohibitions on interference with the work of the Public Defender, as well as the immunities that the office enjoys. In its visiting functions, the PDO

had the power to conduct unannounced visits to closed institutions, a power that is carried over to the NPM, as well as to conduct private interviews with inmates. This confidentiality is explicitly protected against interference. The one weakness in the law, which precludes a maximum score on ATEMPT, is that the protections and immunities that attach to the Public Defender, and to a lesser extent the members of the Special Preventive Group, are not available for those whom the PDO/NPM interviews. Article 21 of the OPCAT explicitly requires that “No authority or official shall order, apply, permit or tolerate any sanction against any person” who communicates with the NPM, and there are clear prohibitions in the Organic Law on the Public Defender (Articles 4 and 5) on interfering or impeding the activities of the Public Defender. However, there is no explicit protection against reprisals for those cooperating with the NPM, in the event either that the clear right of confidentiality is breached or that the authorities simply take measures against those that they presume to be sources of information.

The main difference between the monitoring scores over the ten-year period is not the legal framework but the fact that threats against the staff of the PDO, which had been a feature in the years prior to the establishment of the NPM, have not been repeated since. During that earlier period there were threats against PDO monitors in the Samegrelo region. In 2006, police threatened to arrest and plant drugs on a staff member of the Zugdidi office after he witnessed a detainee being beaten. In 2007, another staff member was accused of “provoking a crime” during a monitoring visit (Carver & Handley, 2016: 416). There has been no recurrence of such threats in recent years.

4.4.1 Overall ATEMPT score

ATEMPT provides not only four individual cluster scores (detention, prosecution, complaints and monitoring) but an overall score that combines the four categories, weighting the more important categories in preventing torture more heavily than the less important categories. A comparison of the overall scores for Georgia in 2008 and 2018 can be found in Table 2. As the scores in this table indicate,

Georgia has implemented changes that have improved their overall torture prevention scores, from 63.47% to 67.62%, but there is much room for improvement. Detention practices remain a problem, and prosecution practices remain a bigger problem since there have been no improvements in the latter at all in the past 10 years.

[Table 2 about here](#)

The ultimate measure of an NPM's effectiveness in the long term is whether the incidence of torture and ill-treatment has reduced. Based on our interviews, the consensus seems to be that torture and ill-treatment in Georgia has declined. Big steps had been made to reduce torture in police and pre-trial custody before the creation of the NPM, but in the past decade the significant improvements have been made in conditions in penitentiaries. Although the NPM had persistently drawn attention to torture and ill-treatment in prisons prior to 2012, it was the emergence of torture as a national political issue that created the momentum for change. Broader improvements in prison conditions have been consequent upon recommendations of the NPM. The move that has had the greatest impact has been the dramatic reduction in the prison population, through amnesties and reformed sentencing policies. This in turn has allowed other systemic improvements, including closure and refurbishment of prisons. The temporary detention isolator system, introduced before the creation of the NPM, remains a highly effective protection in the early days in custody. Although police treatment of suspects remains a great improvement on 15 years ago, there have been persistent reports of excessive force being used in arrests and outside formal custody. This is an issue that the NPM has identified. The situation in psychiatric hospitals is where least progress has been seen in the past decade. The NPM recognizes that it should be a priority to develop capacity and understanding of mental health issues (both in specialized institutions and in prisons).

The architecture of protection was largely in place a decade ago, although there have been marginal improvements, identified by ATEMPT, in detention and monitoring practice. Investigation and prosecution of torturers remains a significant weak point – again one that has been persistently identified by the NPM.

5. Conclusion

This article describes a model for evaluating NPMs that relies in large part on an assessment tool (ATEMPT) that was derived not from a normative indicator framework but from the results of empirical, quantitative research. The model also includes an effectiveness framework (RIPE) that evaluates the formal attributes of the NPM. The model can be supplemented with, if it is available, information obtained from the systematic tracking of NPM recommendations. This combination constitutes a highly effective means of determining the impact of human rights interventions, based upon what is proven to work in practice, rather than upon norms that may be untethered from day-to-day work.

If there is a broader lesson to be derived from this work, it is perhaps that evaluation in the human rights field could be most usefully tied to what works. In other words, we should not only be interested in whether certain inputs lead to predicted outputs, but also whether these outputs result in outcomes, both short- and long-term, that improve human rights. The continued dominance of law within human rights practice has tended to narrow the focus to normative changes that are, ultimately, inputs or, at best outputs. By basing our effectiveness model on rigorously tested data, we were able to prioritize what demonstrably works to prevent torture over what has the strongest normative underpinning, as well as over what practitioners may intuitively ‘feel’ is most important. This may be a particular contribution that scholars have to make to more effective practice, since we are usually at some remove from the quotidian demands of practice. We also have (or should have) a different agenda from the

programme evaluations that many funders require, which may reflect shared assumptions between donors and practitioners about appropriate inputs and outputs. In this instance, our evaluative approach is founded upon data generated from a scholarly project whose sole aim was to discover what works over a long period – on other words which inputs and outputs are most likely to contribute to the desired outcomes.

A related issue is that assessing effectiveness in the human rights field often has to take account of the gap between law and practice, as we did with regard to torture prevention. The finding of our original research was that those prevention measures that have the greatest impact – safeguards in detention and prosecution of torturers – are also those where practice is most likely to fall short of legal standards. For any practitioner, particularly an NPM in this instance, seeking solutions to this shortfall is a priority. This may sound obvious, and in Georgia the NPM did document such shortcomings effectively, but a recent study on European national human rights institutions found that many NHRIs did not monitor or receive complaints relating to violations of detention safeguards because these were perceived to be a matter of law and judicial oversight and hence outside their remit (Monina & Katona 2019). Equally, many NPMs (again not including Georgia) appear reluctant to address the failure to prosecute torturers since this is seen as outside their monitoring role.

We would argue that one of the strengths of our approach is that it goes beyond an evaluation of the work of a specific organization, the national preventive mechanism, and addresses broader national strategies to prevent torture. ATEMPT is a tool designed to assess this entire preventive approach, across the different state actors, and can be used by anyone, whether it be the NPM, executive branch bodies or civil society organizations.

Funding

This work was supported by a grant from the Open Society Georgia Foundation.

Acknowledgments

The authors are grateful for the cooperation and support of the Public Defender's Office, especially Deputy Public Defender Giorgi Burjanadze, Nika Kvaratskhelia, head of the National Preventive Mechanism, and all other NPM staff, with particular thanks to Tamar Khokhobashvili and Akaki Kukhaleishvili.

Bibliography

Birk, M. and others. 2015. *Enhancing impact of national preventive mechanisms*. Vienna: Ludwig Boltzmann Institute for Human Rights and University of Bristol.

Buckland, B. and Csergő, E. 2019. *Strategies and tools for effective follow-up to recommendations*. Geneva: Association for the Prevention of Torture.

Carver, R. 2014. *Measuring the impact and development effectiveness of national human rights institutions: a proposed framework for evaluation*. Bratislava: United Nations Development Programme.

Carver, R. & Handley, L. 2016. *Does Torture Prevention Work?* Liverpool: Liverpool University Press.

Carver, R. & Handley, L. 2019. *Does the Georgian NPM Work? An assessment of 10 years of torture prevention*. Tbilisi: Open Society Georgia Foundation.

Clark, A. M. & Sikkink K. 2013. Information and Human Rights Data: Is the Good News about Increased Human Rights Information Bad News for Human Rights Measures? *Human Rights Quarterly* 35(3): 539-68.

Doran, G. 1981. There's a S.M.A.R.T. Way to Write Management's Goals and Objectives. *Management Review* 70(11): 35-36.

Mertus, J. 2012. Evaluating NHRIs: Considering Structure, Mandate, and Impact. In R. Goodman and T. Pegram (eds), *Human Rights, State Compliance and Social Change: Assessing National Human Rights Institutions*, pp. 74-92 New York: Cambridge University Press.

Monina, G. & Katona, N. 2019. *Strengthening the rights of suspects and accused in criminal proceedings: The role of national human rights institutions*. Vienna: Ludwig Boltzmann Institute for Human Rights.

National Democratic Institute. 2019. *Public Attitudes in Georgia: Results of July 2019 Survey*. Tbilisi:

National Democratic Institute.

United Nations Office of the High Commissioner for Human Rights. 2012. *Human Rights Indicators: A Guide to Measurement and Implementation*. New York and Geneva: OHCHR.

United Nations Subcommittee on Prevention of Torture. No date. NPM assessment matrix for states.

https://www.ohchr.org/Documents/HRBodies/OPCAT/NPM/NPM_Assessment_matrix_for_States.docx

(referenced 9 February 2020)

Yemm, G. 2013. *Essential Guide to Leading Your Team: How to Set Goals, Measure Performance and Reward Talent*. London: Pearson.

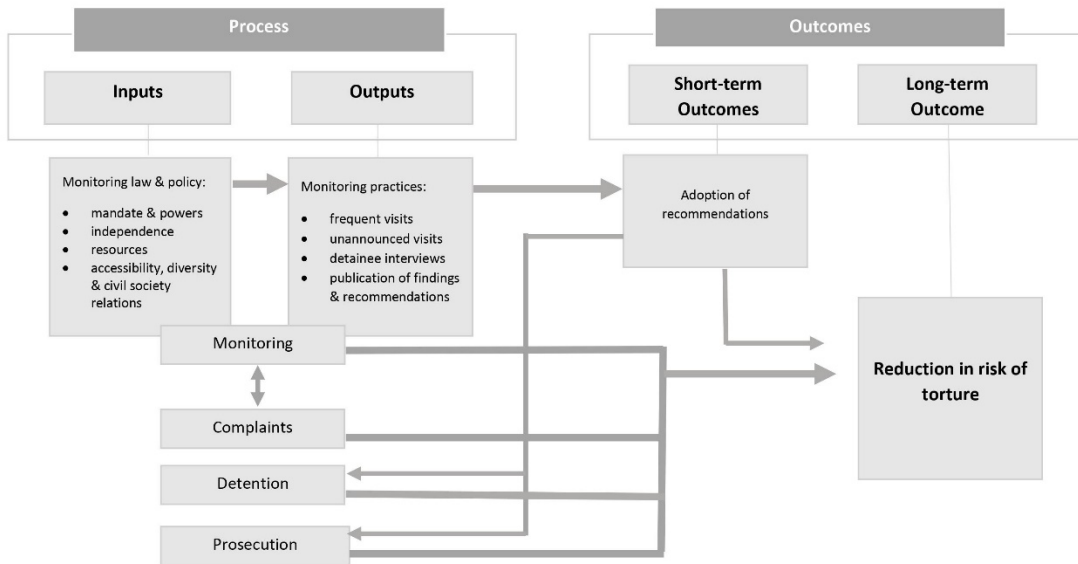


Figure 1: Conceptual model for evaluating NPMs

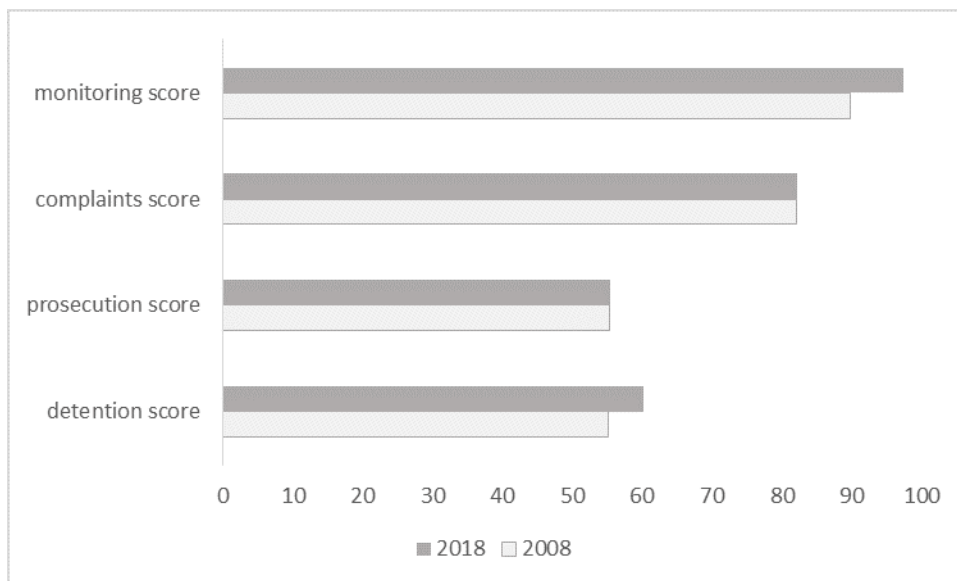


Figure 2: ATEMPT scores, 2008 and 2018

Table 1: Examples of Preventive Measures Included in ATEMPT by Cluster

	Law	Practice
Detention	<ul style="list-style-type: none"> • Unofficial detention illegal • Notification of family required • Right to lawyer • Prompt presentation before judge required • Medical exam required 	<ul style="list-style-type: none"> • Is unofficial detention employed? • Are families notified promptly? • Is right to lawyer conveyed and exercised? • Is presentation before judge prompt?
Prosecution	<ul style="list-style-type: none"> • Criminalization of torture • Substantial penalties for torture • Independent authority to investigate allegations 	<ul style="list-style-type: none"> • Are allegations of torture brought? • Are allegations thoroughly investigated? • Are charges being brought in court? • Are conviction rates comparable to other serious crimes?

<p>Complaints</p>	<ul style="list-style-type: none"> • Independent complaints mechanism • Power to compel evidence and witnesses • Power to refer complaint to investigative authority 	<ul style="list-style-type: none"> • Are complaints investigated effectively? • Are complaints referred to investigative authority (e.g., prosecution)? • Are redress recommendations made?
<p>Monitoring</p>	<ul style="list-style-type: none"> • Domestic monitoring mechanism • Power to make unannounced visits • Power to conduct interviews with detainees 	<ul style="list-style-type: none"> • Does monitor conduct regular and unannounced visits? • Does monitor interview detainees? • Are monitors sanctioned for their activities?

Table 2: ATEMPT Scores by Cluster, 2008 and 2018

Cluster	initial points	weight	final points	Overall Score
2008				
detention	55.13%	0.45	0.248077	
prosecution	55.36%	0.30	0.166071	
complaints	82.14%	0.05	0.041071	
monitoring	89.74%	0.20	0.179487	
SCORE			0.634707	63.47%
2018				
detention	60.26%	0.45	0.271154	
prosecution	55.36%	0.30	0.166071	
complaints	82.14%	0.05	0.041071	
monitoring	97.44%	0.20	0.194872	
SCORE			0.673168	67.32%