Taking the Battle Upstream: Towards a Benchmarking Role for NATO

Stephan De Spiegeleire

Center for Technology and National Security Policy
Institute for National Strategic Studies
National Defense University

September 2012

The views expressed in this paper are those of the author and do not reflect official policy or the position of the U.S. government, the Defense Department, or the National Defense University.

Stephan De Spiegeleire is a Senior Defense Scientist at The Hague Centre for Strategic Studies.

Contents

Executive Summary **1**
Introduction **5**
Defense Benchmarking: A Role for NATO? **7**
NATO's Impact on Capabilities 7
Taking the Battle Upstream 9
Benchmarking – The Origins **10**
Benchmarking Today **11**
Benchmarking – A Working Definition 12
Benchmarking in the Private Sector 13
Benchmarking in the Public Sector 16
Evaluating Benchmarking 20
Fields of Benchmarking *21*
Motives for Benchmarking *22*
Size of the Benchmarking Team *22*
Time Required *23*
Effectiveness of Benchmarking *24*
Benefits of Benchmarking *24*
Benchmarking in Defense Organizations **27**
Defense Benchmarking – The State of the Discipline 27
The 2010 McKinsey Study 28
The Dutch Approach to Defense Benchmarking 33
Background *33*
TNO Report on Defense Benchmarking: A Double Recommendation *34*
TNO Defense Benchmarking Planning Guide *36*
Mainstreaming the Method *37*
Dutch Example of a Defense Benchmarking Study: Capability Development 38
Defense Benchmarking: Where Do We Stand? 53
Conclusion **53**
Bibliography **56**

List of Figures

Figure 1. Moving NATO's Capability Efforts Upstream 1
Figure 2. The Capability Life Cycle 8
Figure 3. NATO's Current Impact on the Capability Life Cycle 8
Figure 4. Taking the Battle for Capabilities 'Upstream' 9
Figure 5. A Surveying 'Bench-Mark' in Cumbria 10
Figure 6. The Use of the Word Benchmarking in 5.2 million Books since 1800 11
Figure 7. A Computing Benchmark 14
Figure 8. World Bank Benchmarking Work on Quality of Governance 17
Figure 9. OECD Benchmarking Work on Education 19
Figure 10. OECD Benchmarking Work on Education 20
Figure 11. Fields of Benchmarking 21
Figure 12. Motives for Benchmarking 22
Figure 13. Typical Size of Benchmark Teams 22
Figure 14. Typical Time for a Benchmark Project 23
Figure 15. Perceived Effectiveness of Benchmarking 24
Figure 16. Future Use of Improvement Techniques 25
Figure 17. Joint vs. Service Spending 31
Figure 18. Tooth-to-Tail Ratio 32
Figure 19. Example of the Topic-to-Metric Decomposition Approach 36
Figure 20. The Main Stages of the TNO Defense Benchmarking Planning Guide 37
Figure 21. The HCSS Audax Index 43
Figure 22. Number of Scenarios Used 44
Figure 23. Specificity of Scenarios 45
Figure 24. The HCSS Z-Chart: Capability Building Process 50
Figure 25. "Nuggets" Distilled from the "Closing the Loop" Benchmark Study 52
Figure 26. Taking the Focus Upstream 54

List of Tables

Table 1. Benefits of Benchmarking 25
Table 2. McKinsey Defense Benchmark: Ratios in 3 Budget Categories 30
Table 3. Active vs. Deployable vs. Deployed Troops 33
Table 4. Benchmarking Ambition Levels in Defense White Papers 40

EXECUTIVE SUMMARY

One of the North Atlantic Treaty Organization's (NATO) goals is to ensure that its member states collectively have the capabilities required to apply decisive force whenever the alliance's political leaders decide to achieve certain effects around the world. Yet the history of NATO's influence on actual defense capabilities is a checkered one at best. Since the height of the Cold War, when NATO set itself a level of ambition of 100 divisions and then promptly proceeded to ignore it, all the way to more recent efforts such as the Defence[1] Capabilities Initiative (1999), the Prague Capabilities Commitment (2002), the Istanbul Usability Targets (2004) and the Lisbon Capabilities Package (2010)—the direct impact of NATO on national capability development has proved disappointing.

If we think of the "life cycle" of defense capabilities from the moment they are conceived to the moment they are disposed of, NATO's effort throughout these years has focused predominantly on the "employment" stage. NATO's Defence Planning Process (NDPP) indicates what its analyses and foresight efforts (and increasingly its operational experiences as well) show is required to be effective in the employment stage and then translates these minimally required capabilities into national targets that are presented to and discussed with the NATO member states. But these collective NDPP inputs remain by and large peripheral to the much more dominant **national** defense planning processes through which the overwhelming majority of Alliance capabilities are "born" and "grown." To put it in business terms: NATO asks for a product and essentially stays aloof from the way(s) in which its providers produce it. In river terms: NATO positions itself '"downstream" where it has to work with the capabilities that the tributaries bring to it. In the NDPP, NATO looks at those contributions and suggests that it would like other capabilities to come downstream, but it does not interfere with the force generation "upstream."

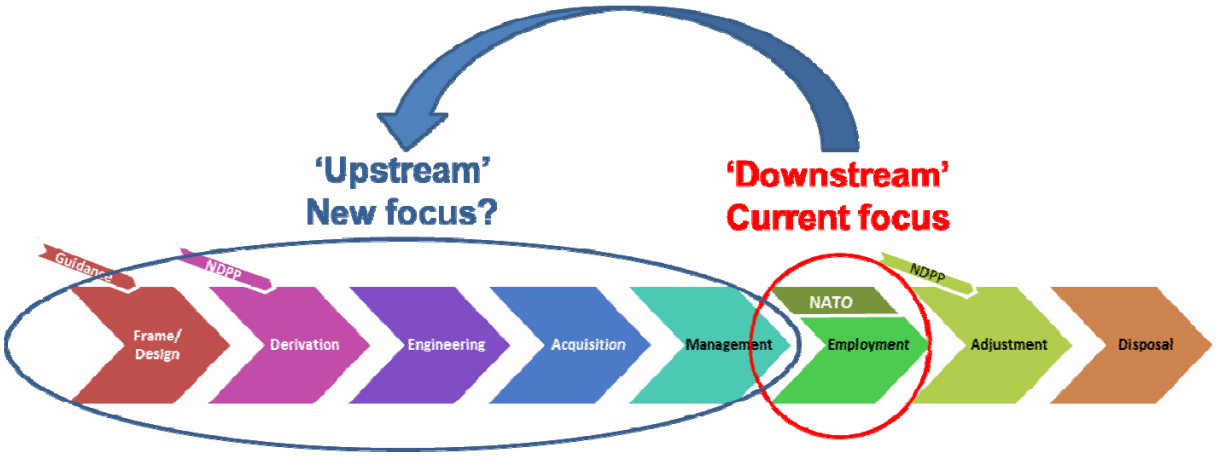

Figure 1. Moving NATO's Capability Efforts Upstream

The main intuition underlying this paper is that the current (geo) political, technological, and especially financial realities may require NATO to take the battle for capabilities upstream. National defense planning processes are one of the most complex planning endeavors on this planet and all NATO nations—even the bigger ones—are struggling with it. There is ample room for improvement through learning from others throughout the capability life cycle. As

[1] In line with NATO practice, this paper will use the British spelling of the word 'defense' whenever it deals with NATO-specific terms, and the U.S. spelling elsewhere.

an international organization, NATO may be ideally placed to facilitate this learning process. At every step in the chevron-chart depicted in Figure 1, each single country makes myriad decisions—big and small—that determine its national force. This force then becomes the pool from which that nation apportions forces to NATO (and not the other way around). Many of these national choices are currently not systematically mapped by any national or international instance. This paper argues that every individual country and the alliance as a whole would greatly benefit from more systematic comparative insights into what works and what does not work in the upstream capability development and management stages.

All nations have to accommodate a large number of diverse (national) pressures in their defense planning efforts: not only operational, but also financial, political, bureaucratic, industrial, employment, and regional. These powerful forces more often than not overwhelm sound analysis, again in large and small Allies alike. This is where cooperative "benchmarking"—also of upstream defense planning processes—might play a uniquely beneficial role: by helping member states to improve the national processes through which capabilities are born and grown or at least to contemplate other solutions than the ones they may come up with in their own capability development and management process.

The bulk of this paper is written as a "primer" in defense benchmarking. Benchmarking remains a relative unknown in the defense arena, despite that fact that it is a technique that is increasingly used in both the private and the public sectors to improve organizational performance through learning from others. This paper defines benchmarking as "an evidence-based analytical effort to systematically compare the products, services, or processes of an organization against those of other organizations in order to improve performance." It differentiates between two different types of benchmarking: benchmarking as a "beauty contest" (normative benchmarking) and benchmarking as "mapping differences" (descriptive benchmarking). *Normative benchmarking* aims to find out which organization does things better or best and typically ends up with some sort of "report card." This form of benchmarking can be extremely effective if, and only if, reliable and widely accepted metrics of performance or effectiveness are available. And even then beauty contests tend to trigger great sensitivities (and resistance) in the organizations that are being benchmarked—often to the detriment of the quality or especially the usefulness of the benchmarking exercise itself. The second, *descriptive* form of *benchmarking* simply sets out to systematically map differences in the ways in which organizations approach various issues and the consequences to which this leads. Especially for more "wicked" problems where there is often not a demonstrably better solution, such a dispassionate mapping exercise can inject more concrete evidence in the decisionmaking process of an organization that is contemplating changes in the way it approaches certain challenges.

Benchmarking has now been used in the private sector for about a quarter of a century. Over this period it has become a standard technique in the strategic management toolkit of many companies. There also is a fairly robust consensus that the practice of benchmarking has helped the organizations that have applied it in their quest to remain competitive. In the public sector, benchmarking started mushrooming about a decade ago and is now widely acknowledged as having assisted "policy transfer" and "policy learning" across countries. Today, many public sector organizations—ranging from central and regional government agencies to police forces and hospitals—are engaged in benchmarking projects that are explicitly aimed at performance improvement. This paper pays special attention to the role international organizations are increasingly playing in this process. It gives some powerful examples from the Organization for Economic Co-operation and Development (OECD), which does much benchmarking work in important policy areas as diverse as education,

health, or innovation policy. The OECD regularly produces and publishes rigorous analyses of the ways in which its member states tackle certain policy issues and the results they achieve. Politicians and policymakers across the world anxiously await these analyses to see how well they score on them and to find out whether there are any other promising approaches from other countries they could adopt.

Defense runs behind on these trends. Defense organizations certainly do often compare themselves to others in an effort to learn. But until recently they have not done so very formally or systematically. A survey of more than 200 defense benchmarking studies showed that defense organizations pay much more lip service to benchmarking than actually engaging in it in a structural, systematic way. Most benchmarking studies tend to be fairly quick and dirty, often based on casual exchanges with other defense organizations, questionable questionnaires, or "benchmarking tourism." On the upside, the survey also found an upward trend in the quantity of explicit defense benchmarks and a few good examples.

This paper showcases what we see as two best-of-kind examples of contemporary defense benchmarks. The **first** example is the large study that the international consultancy McKinsey completed in 2010 in which it compared various aspects of the defense efforts of 33 countries representing roughly 90 percent of global defense spending. The data from this study that were made public reveal stunning ranges across these countries on important aspects of defense such as "tooth-to-tail ratios" that vary from 16 percent to 54 percent or the cost of maintenance per unit of military equipment output (a new metric developed for this study) ranging from $2,000 to $104,000. These striking differences suggest that there is much scope for learning between these organizations—even just based on publicly available data.

The **second example** is the systematic use of benchmarking in the Netherlands Defense Organization. The Netherlands developed and validated a generic planning guide for defense benchmarking in 2006 and the leadership of the organization mandated that any new policy initiative that is put forward has to be subjected to a benchmark feasibility study. This obliges decisionmakers at various levels to look outside of the organization before they make any new choices. The method is based on the systematic decomposition of any topic into concrete metrics derived from authoritative written (and again publicly available) documentation from other defense organizations. Contrary to the McKinsey approach, which is of a more normative nature, the approach here is predominantly descriptive. This paper presents a number of examples from a Dutch benchmarking study of the ways in which countries do capability planning. These examples illustrate that benchmarking can often just highlight important differences in approaches that at least force decisionmakers to think about such alternatives (and the possible consequences they may have led to in other countries). On top of executing a growing number of such studies as part of the regular military planning, programming and budgeting system, the Netherlands has also trained about 100 Ministry of Defence staff members (both military and civilian) in the method, and an even larger number has now had first-hand experience with defense benchmarking. A number of these benchmarking studies have also led to different choices than would have been made without this initial "outward" look.

These two very different, but complementary "best of kind" approaches to defense benchmarking demonstrate that there is enough publicly *available* information to arrive at meaningful comparisons that can be used by defense organizations to improve their performance. Defense organizations publish ever larger quantities of information and data to satisfy increasingly more demanding *national* reporting requirements. Much work remains to be done to collate these data—which are currently vastly underused—in a more systematic

way and to make them reliably (and traceably) comparable. But such an effort is likely to be quite beneficial to both individual countries and to the Alliance as a whole.

National efforts (both unilateral and "minilateral") to learn from others in the defense and security area will undoubtedly continue. We also surmise that consultancies will continue to build up and exploit their own proprietary knowledge bases with the comparative insights they glean from the work they do for various defense organizations across the world. Defense organizations are likely to benefit from both of these efforts and it might even be useful to explore ways to come to some form of public-private partnership between these two efforts. But currently we still feel a preferable model would be for some international organization like NATO to assume this task by creating a clearinghouse of evidence-based benchmarking insights to the benefit of its member states—along the lines of the work that the OECD does in other policy areas. Efforts by individual (or small groups of) nations, companies, or think tanks can certainly provide valuable inputs that can be used by decisionmakers across the Alliance (provided they are made publicly available, preferably in English). But they are unlikely to singlehandedly be able to overcome the various hurdles (also analytical) that rigorous defense benchmarking encounter. To be truly effective, defense benchmarking is in need of a higher-level catalyst, a strategic engine. NATO—and particularly its Allied Command Transformation, the Alliance's leading agent for change "driving, facilitating, and advocating continuous improvement of Alliance capabilities to maintain and enhance the military relevance and effectiveness of the Alliance"—is ideally placed for such a role. It has the mandate, the authority, and the resources to build up a more systematic benchmarking facility within the Alliance. The knowledge base such a facility would produce could be put at the benefit of national defense planners, thus taking the battle for better capabilities upstream. In this way, defense benchmarking could become a new tool in a richer and "smarter" strategic defense management toolbox in line with what NATO's new push for "smart defense" is trying to achieve.

INTRODUCTION

> *Lesson-drawing is practical; it is concerned with making policies that can be put into effect. The point of learning is not to pass an examination; lessons are meant to be tools that guide actions. As long as government proceeds routinely policymakers may assume that established policies are satisfactory; the guiding maxim is: 'If it ain't broke, don't fix it'. But what happens when an increase in dissatisfaction creates a demand to do something?*[2]

The area of national defense has always been a reflective one. Throughout history both armed forces and their political-military leaders have gone to great lengths to learn—from themselves, from their predecessors, and from others. This age-old learning instinct (some may call it "stealing" or "spying") is now being boosted throughout the North Atlantic Treaty Organization (NATO) Alliance by some important new challenges and opportunities.

The ***increased use of our armed forces*** in both low- and high-intensity operations over the past few decades has laid bare the glaring differences between NATO countries much more clearly and painfully than any political rhetoric about burdensharing ever could. This has led to frustrations of a number of political and military leaders—both domestically ("why can't we…") and comparatively ("how come they can…"). Similar vexations are sparked by the accelerating ***pace of change*** (technological, organizational, doctrinal, political, etc.) in all spheres of life—including the defense one—making it ever more difficult to "keep up" with "the others," "the private sector," "technological innovation," and the like. Both national and international pressures are ***squeezing defense budgets*** at the very time when politicians across the Alliance are (re)discovering the utility of the military instrument from places like Libya to Afghanistan. This necessitates a much more efficient allocation of scarce resources and a willingness to learn from others in this area.

At the same time, there are also a number of new opportunities for benchmarking that just did not exist before. There is more ***transparency*** today about military affairs than ever before in history—including (and even especially) by the leading military powers—offering unprecedented opportunities to learn even just from what they make available in the public domain. In this increasingly global world, military establishments also ***interact more*** with each other in cooperative ways than ever before; this direct contact is reinforcing the natural trend of defense organizations to learn from others. Lastly, the various taboos that have historically led to the isolation of the military field from other fields of public and private policy are starting to break down and the pressures (and incentives) to ***learn, especially from the private sector,*** are growing.

As a consequence of these changes, the desire to improve defense organizations' value proposition by "learning from the best" is becoming almost irresistible. The emergence of benchmarking (and other related data-driven, evidence-based planning tools) as one of the leading methodologies used in the private sector to improve performance naturally feeds into this burgeoning desire to compare oneself with others and to learn from the best.

---

[2] Rose, "Ten Steps in Learning Lessons from Abroad."

This paper about defense benchmarking is set against this broader background. The immediate trigger for it is the recent push within NATO for smart defense. NATO Secretary General Rasmussen has put great emphasis on this concept by encouraging nations to maintain and improve their capabilities despite the financial crisis by making better use of resources.

> *Smart Defense is about nations building greater security—not with more resources, but with more coordination and coherence.*[3]

Most of the current discussions within the Alliance on smart defense are focused on better forms of multinational "pooling and sharing," but there is also much new thinking on how we can improve NATO defense planning. As part of the new NATO Defence Planning Process and on the basis of the new (public) NATO Strategic Concept that was agreed at the 2010 Lisbon summit, NATO is issuing more detailed (classified) *Political Guidance* for the Alliance's defense planning efforts. This is intended to be a single, unified political guidance for defense planning that sets out the overall aims and objectives to be met by the Alliance. The main part of this document aims at defining the number, scale, and nature of the operations the Alliance should be able to conduct in the future (commonly referred to as NATO's Level of Ambition). The intention here is that this consolidated guidance will steer the capability development efforts of Allies and within NATO.[4] But in another part, the new political guidance document also spells out the need for better defense metrics. The main idea here is to obtain a more comprehensive picture of how and where Allies use their defense resources. These new metrics, which are to cover a range of input and output measurements, are supposed to complement the ones that are currently collected through the NATO Defense Planning Capability Survey (DPCS, formerly known as the Defense Planning Questionnaires or DPQs)[5] and the NATO usability initiative.[6] This clarion call for better metrics was taken up by NATO Allied Command Transformation (ACT) through its Joint Analysis and Lessons Learned Centre (JALLC) in Lisbon, Portugal. JALLC's commander, Brigadier General Peter Sonneby, convened a mixed working group under the lead of Dr. Bent-Erik Bakken from the Norwegian Defense University College to provide an analytical input into the Alliance's discussion about new metrics. The bulk of that effort has been devoted to identifying a new set of possible defense metrics that could complement and add value to the already existing set of metrics in order to start providing the "more comprehensive picture" the Alliance is looking for. But at the same time, The Hague Centre for Strategic Studies (HCSS) was also tasked by NATO JALLC to provide an additional reflection paper on the concept and practice of benchmarking in the defense area.

This paper represents the HCSS contribution to this debate. It is conceived as a primer in defense benchmarking and is structured in five sections. The ***first*** section presents the main argument of the paper: that NATO should take the battle for better capabilities upstream to

---

[3] "NATO - Opinion: NATO – Value for Security'' - Speech by NATO Secretary General Anders Fogh Rasmussen in Bratislava, Slovakia, May 19, 2011.

[4] "NATO - Topic: Defence Planning Process."

[5] Ibid.

[6] Usability goals for land forces personnel—that 40 percent of should be deployable and 8 percent sustainable—were established at the 2004 NATO Summit in Istanbul. In 2008 and 2009, the targets were raised to 50 percent and 10 percent respectively. In 2010, Allies agreed air usability targets, based on the counting of airframes, that 40 percent should be deployable and 8 percent sustainable.

the heart of the national (forward) defense planning processes. The rest of the paper is written as a primer on defense benchmarking. The ***second*** and ***third*** sections of the paper briefly discuss where the concept of benchmarking came from and where it stands today—both in the private sector and the public sector. In the ***fourth*** section we turn our attention to defense benchmarking proper. This section starts with a "state of the discipline" overview and then devotes special attention to two notable examples of defense benchmarking: the institutionalized practice of benchmarking in the Netherlands Defense Organization (as an example of more descriptive benchmarking that essentially tries to map differences without making judgment calls) and the 2010 McKinsey defense benchmark (as an example of more normative benchmarking that tries to discover which country does better or worse on some key aspects of defense). This section wraps up with some concrete examples of recent benchmarking work in an area related to the broader topic of the paper: how countries derive and develop their defense capabilities. The paper concludes in the ***fifth*** section with some final reflections about the need for a higher-level catalyst for rigorous defense benchmarking and the role NATO ACT could play in this.

DEFENSE BENCHMARKING: A ROLE FOR NATO?

NATO's Impact on Capabilities

It is one of NATO's ambitions to ensure its member states collectively have the capabilities required to apply decisive force whenever the alliance's political leaders decide to use NATO to achieve certain effects across the world. Currently, much of the Alliance's efforts are quite understandably focused on ongoing operations. That implies that political and military leaders have to plan operations with the existing capabilities that Allies are willing to allocate to NATO. At the same time, however, the Alliance also works on future capabilities through the (recently reworked) NDPP, in which it strives to make sure Allies have the necessary capabilities required to cover all missions that political leaders have entrusted upon the organization. In order to do so, it derives a set of minimum capability requirements (including shortfalls, where applicable) from the politically approved mission set and then apportions those to nations.

In reality, the history of NATO's influence on actual capabilities is a checkered one at best. Since the height of the Cold War during the Korean War, when NATO set itself a level of ambition of 100 divisions (at a time when NATO's entire posture still numbered 12 divisions) and then promptly proceeded to ignore it, all the way to more recent efforts such as the Defence Capabilities Initiative (1999), the Prague Capabilities Commitment (2002), the Istanbul Usability Targets (2004) and now the Lisbon Capabilities Package (2010)—the impact of NATO on national capability development has been disappointing.[7] Capabilities typically mean money and NATO allies have always been reluctant to "socialize" defense capabilities meaning the money—and the capabilities—remain fiercely national. The only NATO-owned and operated capabilities at this moment are the NATO's Airborne Early Warning and Control (NAEW&C)—also known as AWACS—radar aircraft. All other Alliance capabilities are born and grown nationally in national processes over which NATO has little to no influence. Figure 2 tries to map the generic life cycle of a capability.

[7] Kugler, *Laying the Foundations*, 56 ff.

Figure 2. The Capability Life Cycle

The first step in this scheme is one we have called the "design" (or framing) stage of capability planning. It is a step that is often overlooked, but the way in which we conceive of capabilities greatly affects the actual capabilities we obtain.[8] Within this particular capability frame, we then proceed to define the actual concrete capabilities that are thought to be required to fulfill the scope of ambitions of the political leadership. Since the introduction of capability-based planning in the past decade, this derivation process in many (especially larger) countries (and in NATO itself) now typically translates political guidance to capabilities by using a set of scenarios that are thought to be representative for the operations in which armed forces might get involved.[9] In many smaller countries, this process tends to be less formalized and more "marginal" in the sense that it focuses mainly on changes to the existing force that are imposed by the environment or—even more frequently—by funding cuts or by the obsolescence of certain existing capabilities.

As soon as new capabilities are defined they either have to be "engineered" in case they do not yet exist or acquired in case they do. Once engineered and acquired they enter the armed forces to be maintained at certain levels of readiness and—when and where required—employed. After such employment, they often have to be adjusted on the basis of altered requirements or new possibilities. At the end of their life cycle, they also have to be disposed of—another part of the life cycle that is not typically thought of but can be quite consequential.

Figure 3. NATO's Current Impact on the Capability Life Cycle

Figure 3 visualizes our own view of where NATO currently impacts what remains essentially a national process. The bulk of that impact, as we pointed out, is focused on the employment part of the life cycle—what we will call the downstream of the process (the right side of the chevron-diagram in Figure 3). When NATO embarks on a military operation, the slice of national capabilities that countries pledge to that operation for all intents and purpose really does become "NATO." NATO's impact on the other parts of the capability life cycle, however, is much more modest and mostly indirect. NATO strategic guidance (contained in documents such as the Strategic Concept or the Comprehensive Political Guidance) is mostly intended for the Alliance as a whole, but could be said to have a certain impact on the way in

[8] We have argued elsewhere that our current conception is one that remains firmly embedded in the industrial age. De Spiegeleire, *Defence Planning*.

[9] De Spiegeleire et al., *Closing the Loop. Towards Strategic Defence Management*.

which countries frame/design their capabilities. NDPP also clearly plays some role in at least some countries—by all evidence much more so in the "new" NATO members than in the "old" ones—through the targets that are apportioned to them and thus become an input (alongside many other ones) in the national capability derivation and adjustment stages of the life cycle. There are a number of additional areas where NATO also has some impact on national processes but as Figure 3 suggests, the overwhelming majority of steps in this process remain national until capabilities are actually employed. To put it somewhat cynically: whenever the outcomes of the NDPP happen to coincide with this (dominant) national process – capabilities are generally delivered. Whenever they do not, the experience of the past few decades shows that NATO targets are unlikely to be met.

Summing up, NATO's efforts throughout these years have focused predominantly on the employment stage to the right (downstream) side of the chevron-diagram. NDPP identifies what its analyses and foresight work (and increasingly also its operational experiences) show is required to be effective in the employment stage and then translates these minimal required capabilities into national targets that are presented and discussed with nations. But these NDPP inputs remain by and large external to the much more dominant national defense planning processes through which overwhelming majority of Alliance capabilities are born and grown. To put it in business terms: NATO asks for a product, and essentially stays aloof from the way(s) in which this product is produced by its providers. To put it in more poetic terms, NATO positions itself downstream of the "river" where it has to work with the capabilities that the various tributaries to the river bring to it. In the NDPP it looks at those and sends signals that it would like other capabilities to come downstream, but it does not interfere directly with the upstream.

Taking the Battle Upstream

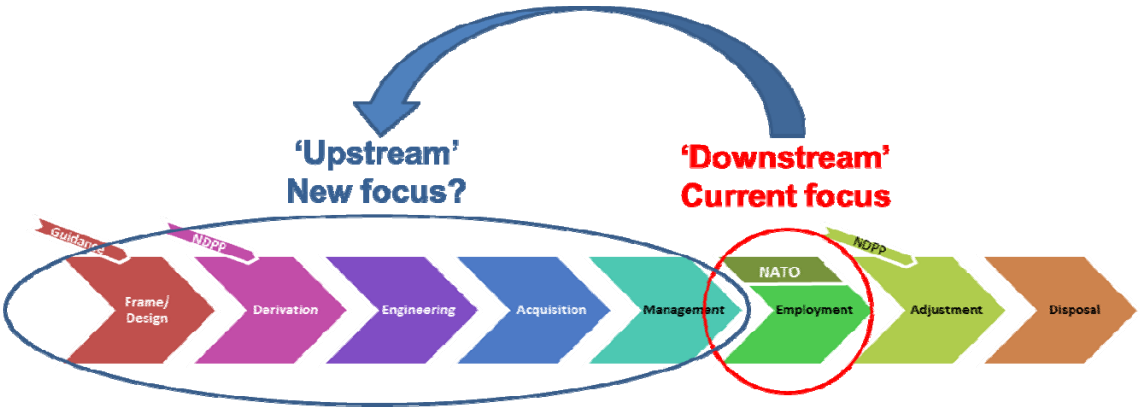

Figure 4. Taking the Battle for Capabilities 'Upstream'

One of the main intuitions underlying this paper is that there is ample room for improvement —and for learning from each other—throughout the capability lifecycle. At every step in this chevron-chart each individual country makes myriad decisions—big and small—that affect the ultimate force that becomes the pool from which countries apportion forces to NATO (and not the other way around). Many of these choices are currently not systematically mapped by any national or international instance. Yet, as Figure 4 suggests, every country, and the alliance as a whole, could greatly benefit from more comparative insights into what works and what does not work in the upstream capability development and management stages. Managing the life cycle of defense capabilities is indeed a Herculean task with which all countries struggle. All have to accommodate a large number of diverse (national) perspectives: not only operational, but also financial, political, bureaucratic, industrial, and

employment. Confronted with all these powerful forces, sound analysis more often than not suffers. This is where cooperative benchmarking might be able to play a role: by helping member states in at least contemplating other solutions than the ones they may come up with in their own capability development and management process.

BENCHMARKING – THE ORIGINS

The word "benchmark" has become part of the everyday vocabulary in many fields. And yet the background of this word is not widely known and may therefore deserve some attention, all the more since few people realize the term actually originated in a military context.

The meanings of both components of this word—"bench" and "mark"—are quite well known. A bench is something one can sit on, and a mark is a visible trace or sign. But the combination of these two words remains somewhat puzzling—even to native speakers. To unravel this puzzle we have to go back to the military history of England in the mid-18th to early 19th century.[11] In this period England was confronted with a number of serious military challenges both in the North, with continued unrest in the Scottish Highlands after the Jacobite Rising of 1745, and in the South, where an ascendant France was viewed as a growing territorial threat to the British Isles. It was in this context that King George II decided to embark upon a military survey of the entire country. The intent here was that higher-quality data, in this case geographical data, would give England a comparative military advantage over its potential enemies. This resulted in the Principal Triangulation of Great Britain (1783–1853) and the creation of the Ordnance Survey, which was a branch of the British armed forces at that time. The whole triangulation effort required identifying "fixed" points (often on churches) of known elevation that could be used to start measuring the elevation of various other objects across the country. The land surveyors who carried out this effort started chiseling horizontal *marks* throughout the country to mark points of known vertical elevation.

Figure 5. A Surveying 'Bench-Mark' in Cumbria[10]

As Figure 5 shows, these *marks* were usually highlighted with a chiseled arrow below a horizontal line that was also carved out in a stone. This allowed military land surveyors to place an angle-iron in those marks to bracket (*bench*) a leveling rod, thus ensuring that the leveling rod could be repositioned in the same place in the future. This allowed subsequent surveyors to establish the elevation of nearby points through triangulation. A benchmark is thus in essence a fixed point of reference of which the elevation is known or assumed and that can be used to determine the elevation of other objects.[12]

---

[10] "Lakes Guides, Bench Marks, Cumbria, Frameset."

[11] Seymour, *A History of the Ordnance Survey*; Hewitt, *Map of a Nation*.

[12] Venkatramaiah, *Textbook of Surveying*, 123.

It is important to point out that there was nothing normative about the original meaning of the word benchmark. A higher benchmark was not better than lower one or vice versa. A benchmark also was not a target to be aspired to. It was merely a metric that allowed to rigorously compare one data point with another, to get a comprehensive picture of the entire landscape.

BENCHMARKING TODAY

From its origins in land surveying, the concept of benchmarking branched out in a number of different directions. Today benchmarking is "in." The term is used with increasing frequency in a growing variety of areas – as illustrated in Figure 6 that plots how often the word benchmarking appeared in the 5.2 million books published in the past two centuries that Google was able to digitize to date.[13]

Figure 6. The Use of the Word Benchmarking in 5.2 million Books since 1800

In the business world, benchmarking became a standard management tool in the 1990s around which an entire cottage industry of consultants has since mushroomed. The trend took some years to spill over into the public sector, but also here benchmark studies are currently being performed on issues ranging from public corruption to educational quality. Today the word benchmark even emerges in unexpected contexts as when the United States issued benchmarks for the Iraqi government—a set of 18 (congressionally mandated) political and security criteria the Iraqi government had to live up to.[14] In this part of this paper, we will

[13] This represents roughly 4 percent of all books ever published. For more details see Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books"; Bohannon, "Google Books, Wikipedia, and the Future of Culturomics." The web-based interface to this corpus is available at <http://ngrams.googlelabs.com>.

[14] Katzman and Congressional Research Service, *Iraq*.

first provide a generic definition of the term benchmark and will then proceed with a quick overview of some of the main applications of benchmarking in the defense and the non-defense sectors.

Benchmarking – A Working Definition

It may be useful to provide a working definition of the term benchmarking. As with so many terms, there is a vigorous debate in the academic community about what benchmarking actually means.[15] One study even identified 49 definitions for benchmarking,[16] with the differences mainly due to slightly different views on issues such as formality, metrics, comparability, descriptive vs. normative, and linkages with implementation and organizational improvement.[17] Still the fundamental ideas behind benchmarking are broadly shared and can in our view be summarized in the following three main components:

- to compare certain aspects (products, services, or processes) of one's organization with those of other organizations (the **comparative** component)
- based on systemically comparable data (the **data-driven** component)
- with the aim of improving one's performance (the **performance-enhancing** component).[18]

We therefore propose the following generic working definition for the term benchmarking: "an evidence-based analytical effort to systematically compare the products, services, or processes of an organization against those of other organizations in order to improve performance."[19]

We want to emphasize that this broad definition takes out the frequently encountered normative component by which benchmarking quickly transforms in what could be called a "beauty contest."[20] We already showed that the original meaning of the word was not normative in nature, but merely descriptive. But more importantly, we see this broader definition as a more pragmatic approach to the ongoing debate about benchmarking as a beauty contest vs. benchmarking as "mapping differences"—also (but not exclusively) in defense planning. Our own take on this is that wherever it is possible to make well-founded and validated normative judgments, organizations are well advised to pursue and heed them. We feel, however, this is only possible in areas where reliable measures of effectiveness are available on which to base such judgments. In those cases—and only in those cases—can

---

[15] Talluri and Sarkis, "A Computational Geometry Approach for Benchmarking"; Nandi and Banwet, "Benchmarking for World Class Manufacturing–concept, Framework and Applications"; Anand and Kodali, "Benchmarking the Benchmarking Models"; Anderson and McAdam, "Reconceptualising Benchmarking Development in UK Organisations."

[16] Nandi and Banwet, "Benchmarking for World Class Manufacturing–concept, Framework and Applications."

[17] Anand and Kodali, "Benchmarking the Benchmarking Models."

[18] See also Anderson and McAdam, "Reconceptualising Benchmarking Development in UK Organisations."

[19] This comes close to the U.S. Army definition of benchmarking: "a systematic process of comparing, measuring, and analyzing the products, services, or processes of an organization against current best practices of other (preferably world-class) organizations in order to attain superior performance."

[20] Already a 1999 article on benchmarking in the public sector warned against this: "the best benchmarkers resist the tendency for benchmarking to become a beauty contest. It is a powerful tendency, the quest to claim the number one ranking and, perhaps more significantly, to avoid the embarrassment of an unfavorable rank." Ammons, "A Proper Mentality for Benchmarking," 108.

differences in techniques, choices, or approaches be gauged against the observable quality of their effectiveness or performance.

But for more "wicked" problems where such reliable measures are not available or are hotly contested (and there are very many of those in the defense realm), we submit that systematic comparisons can still help the strategic planning and management efforts of an organization.[21] This holds all the more true in periods of rapid complex change in which success may prove fickle and in which a rich portfolio of strategic "experiments" that can adaptively be augmented or scaled down based on changing circumstances may hold the key to long-term success.[22] In this case, knowing and tracking the strategic choices others have made might help an organization—and *a fortiori* an alliance—in navigating turbulent waters. It may not be obvious whether one option is better or worse than another, but being aware of the different options available to both oneself and to others (and their outcomes) enriches evolutionary learning opportunities.

Benchmarking in the Private Sector[23]

One of the best ways to illustrate the essence of benchmarking is to refer to an area that many of us are probably familiar with: the **computer world**. When a consumer wants to buy a new computer, there are a number of standard benchmarking tools (many of them embedded in software programs) that can assist in assessing the relative performance of an object by running a set of standardized tests and trials against it.

They thus provide a method of comparing the performance of various subsystems across different chip/system architectures—often (but not always) with reliable performance metrics. Popular computer magazines and websites frequently feature such benchmarks in their reviews of soft- or hardware. Figure 7 depicts a recent benchmark of how network use

---

[21] 'Wicked problems' are problems that are hard or impossible to solve because of incomplete, contradictory, and changing requirements that are often difficult to recognize. Moreover, because of complex interdependencies, the effort to solve one aspect of a wicked problem may reveal or create other problems. For the seminal formulation of this problem, see Rittel and Webber, "Dilemmas in a General Theory of Planning."

[22] See the 'Red Queen' chapter of Beinhocker, *The Origin of Wealth*.

[23] For those more interested in the literature on this topic, we recommend the following reading list: Adebanjo, Abbas, and Mann, "An Investigation of the Adoption and Implementation of Benchmarking"; Adebanjo et al., "Twenty-five Years Later–a Global Survey of the Adoption and Implementation of Benchmarking"; Adebanjo, Mann, and Abbas, "Benchmarking - BPIR.com"; Adebanjo, Abbas, and Mann, "An Investigation of the Adoption and Implementation of Benchmarking"; Ahmed and Rafiq, "Integrated Benchmarking"; Anand and Kodali, "Benchmarking the Benchmarking Models"; Andersen and Pettersen, *The Benchmarking Handbook*; Anderson and McAdam, "An Empirical Analysis of Lead Benchmarking and Performance Measurement"; Anderson and McAdam, "Reconceptualising Benchmarking Development in UK Organisations"; Auluck, "Benchmarking"; Camp, *Benchmarking*; Dattakumar and Jagadeesh, "A Review of Literature on Benchmarking"; Fernandez, McCarthy, and Rakotobe-Joel, "An Evolutionary Approach to Benchmarking"; Fong, Cheng, and Ho, "Benchmarking"; Francis and Holloway, "What Have We Learned?"; Hinton, Francis, and Holloway, "Best Practice Benchmarking in the UK"; Ginn and Zairi, "Best Practice QFD Application"; Kyrö, "Revising the Concept and Forms of Benchmarking"; McCarthy and Tsinopoulos, "Strategies for Agility"; Moffett, Anderson-Gillespie, and McAdam, "Benchmarking and Performance Measurement"; Moriarty, "A Theory of Benchmarking"; Nandi and Banwet, "Benchmarking for World Class Manufacturing–concept, Framework and Applications"; Papaioannou, Rush, and Bessant, "Benchmarking as a Policy-making Tool"; Raa, *The Economics of Benchmarking*; Zairi and Léonard, *Practical Benchmarking*; Voss, Åhlström, and Blackmon, "Benchmarking and Operational Performance"; Zairi, *Effective Benchmarking*; Zairi, *Effective Management of Benchmarking Projects*; Zairi, *Benchmarking for Best Practice*.

affects the computer's central processing unit across a number of new motherboards. We observe that in this case, it is possible to make a normative assessment: lower use is better.

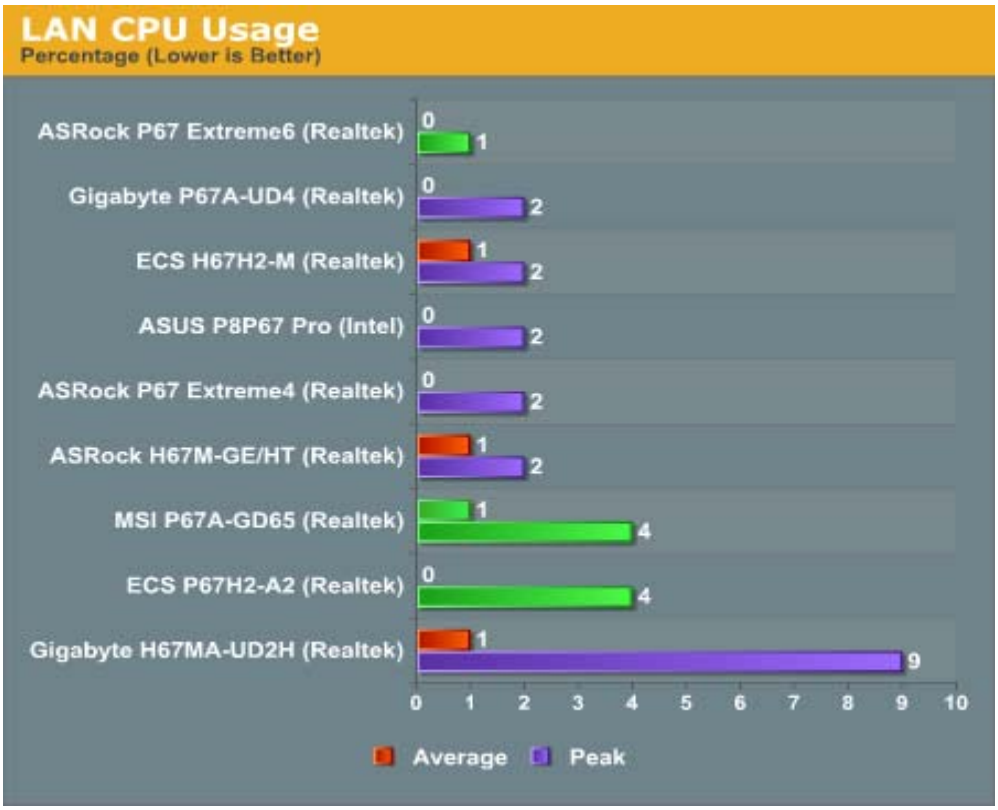

Figure 7. A Computing Benchmark

In the world of "hard" technology—of which there are clearly many examples in the defense world as well—such "hard" benchmark studies are quite common (i.e., with reliable, validated, and widely accepted quantitative metrics on both the parameters of the item to be benchmarked and the output of those parameters).

But also in the **business world,** a "softer" version of benchmarking has become a standard tool in performance management. The business benchmarking methodology was pioneered in the late 1980s by Robert C. Camp at Xerox.[24] Up to that point, companies often tried to learn from their competitors, but they did so primarily by focusing on the finished products and then relying on "reverse engineering" those in order to unravel product design clues. Xerox, however, started taking a much closer and more systematic look not just at the products themselves ("output"), but also at the different manufacturing and other supporting processes that produced them ("throughput"). In the mid-1970s, Fuji-Xerox, Xerox' Japanese joint venture with Fuji photo, and other Japanese competitors started manufacturing experimental copiers at significantly lower costs than U.S.-based Xerox. As this started threatening Xerox' leading market position Xerox CEO David Kearns and Robert Camp, the logistics engineer who initiated Xerox's benchmarking program, set out to systematically analyze Japanese manufacturing costs and product design differences compared to their own. By studying and then adopting/adapting these Japanese companies' demonstrably superior manufacturing

[24] Camp, *Benchmarking*.

(hard) and business (softer) processes Xerox was able to cut average manufacturing by 20 percent and the time-to-market for new products by 60 percent.[25]

These impressive figures (and Camp's subsequent book about this experience[26]) garnered much attention and led to development of an entire cottage industry around benchmarking. Already in 1999, 10 years after the publication of Camp's book, a survey identified benchmarking as one of the top five management tools.[27] Since then, benchmarking has become a formally recognized criterion in a number of quality management standards such as the U.S. National Institute of Standards and Technology Baldrige criteria for Performance Excellence used for the Baldrige Award (an annual award given by the U.S. Department of Commerce to a small set of organizations which demonstrate excellence in quality);[28] the EFQM's (formerly known as the European Foundation for Quality Management) Excellence Model,[29]; and the "Total Quality Management" principles (including the International Standards Organization's ISO 9000 family of quality standards).[30] The data-driven, methodical approach of another popular management approach called Six Sigma also dovetails nicely with benchmarking.[31] There are professional associations for benchmarking practitioners such as the Strategic Planning Institute's Benchmarking Council as well as organizations that act as clearinghouses for benchmarking information and benchmarking case studies (e.g., the International Benchmarking Clearinghouse sponsored by the American Productivity and Quality Center [APQC]).[32]

---

[25] We want to emphasize some interesting analogies between this schoolbook example of private sector benchmarking and defense benchmarking within NATO. Fuji-Xerox was a member of the Xerox "alliance" that just did certain things differently (and—in a number of cases—demonstrably better) than Xerox-US. By looking for the right metrics on both input, especially throughput **and** output, Xerox' CEO David Kearns was able to adopt (in some cases, adapt) what his team felt were superior techniques. Just as in Xerox case, NATO also has a number of allies in its alliance (and its ecosystem) that do things differently. The cooperative form of (intra-Alliance) benchmarking that lead Xerox to such startling improvement results (and has done the same for numerous other companies since then [e.g., Google's permanent internal CD&E efforts]) may therefore lead to improvements in defense management in a number of countries, to the benefit of those countries themselves **and** of the Alliance as a whole. Another fascinating point is the story behind Fuji-Xerox and the advantages (**and** disadvantages) that Xerox' internal diversity (Fuji Xerox owned some assets and Xerox owned others; Fuji Xerox had rights to the Japanese market and Xerox to the United States market; Xerox did not have full control over the capabilities of Fuji Xerox, even though it owned part of the venture's equity) gave it over its more monolithic main competitor Canon. See Gomes-Casseres, "Competing in Constellations."

[26] Camp, *Benchmarking*.

[27] Wong and Wong, "A Review on Benchmarking of Supply Chain Performance Measures."

[28] "Baldrige Criteria for Performance Excellence." – see N2.

[29] EFQM even authored an interesting European Benchmarking Code of Conduct (European Foundation for Quality Management, "European Benchmarking Code of Conduct.")

[30] It notably also cautions (as do we) against participating in any "benchmarking activity that is nothing more than industrial tourism and/or copying. The first step in benchmarking, if undertaken, should be to understand the "what and why" of current performance of your own system or process. That work usually exposes substantial scope for action for improvement." Hoyle, *ISO 9000*, 15.

[31] Six Sigma is a quality management initiative that aims to eliminate defects to reach six standard deviations from the desired target of quality. Six standard deviations means 3.4 defects per million. On benchmarking and Six Sigma, see Watson, *Strategic Benchmarking Reloaded with Six Sigma*.

[32] "Open Standards Benchmarking Assessments - APQC."

Benchmarking in the Public Sector

The practice of benchmarking also engulfed the public sector in the mid-1990s with Europe (and especially the United Kingdom) in a leading role.[33] The European Union (EU) has continued to play a big role in the systematic comparison of various policy areas through its 'open method of co-ordination' with its focus on the identification and dissemination of 'best practice' through mutual learning and peer review, offering new solutions for policy management in an increasingly complex, diverse and uncertain environment.[34]

Today, many public sector organizations—ranging from central and regional government departments to police forces and hospitals—are engaged in benchmarking projects that are aimed explicitly at performance improvement.[35]

---

[33] Bowerman et al., "The Evolution of Benchmarking in UK Local Authorities."

[34] Room, "Policy Benchmarking in the European Union." See also the EU's "European Benchmarking Network."

[35] Braadbaart and Yusnandarshah, "Public Sector Benchmarking"; Bullivant, *Benchmarking for Continuous Improvement in the Public Sector*; Cowper and Samuels, "Performance Benchmarking in the Public Sector"; Dorsch and Yasin, "A Framework for Benchmarking in the Public Sector"; Flynn, *Public Sector Management*; Hood, Dixon, and Beeston, "Rating the Rankings"; Jarrar and Schiuma, "Measuring Performance in the Public Sector"; Lundvall and Tomlinson, "International Benchmarking as a Policy Learning Tool"; Magd and Curry, "Benchmarking"; Triantafillou, "Benchmarking in the Public Sector"; Tillema, "Public Sector Benchmarking and Performance Improvement"; ibid.; Tillema, "Public Sector Organizations' Use of Benchmarking Information for Performance Improvement."; Van Helden and Tillema, "In Search of a Benchmarking Theory for the Public Sector."

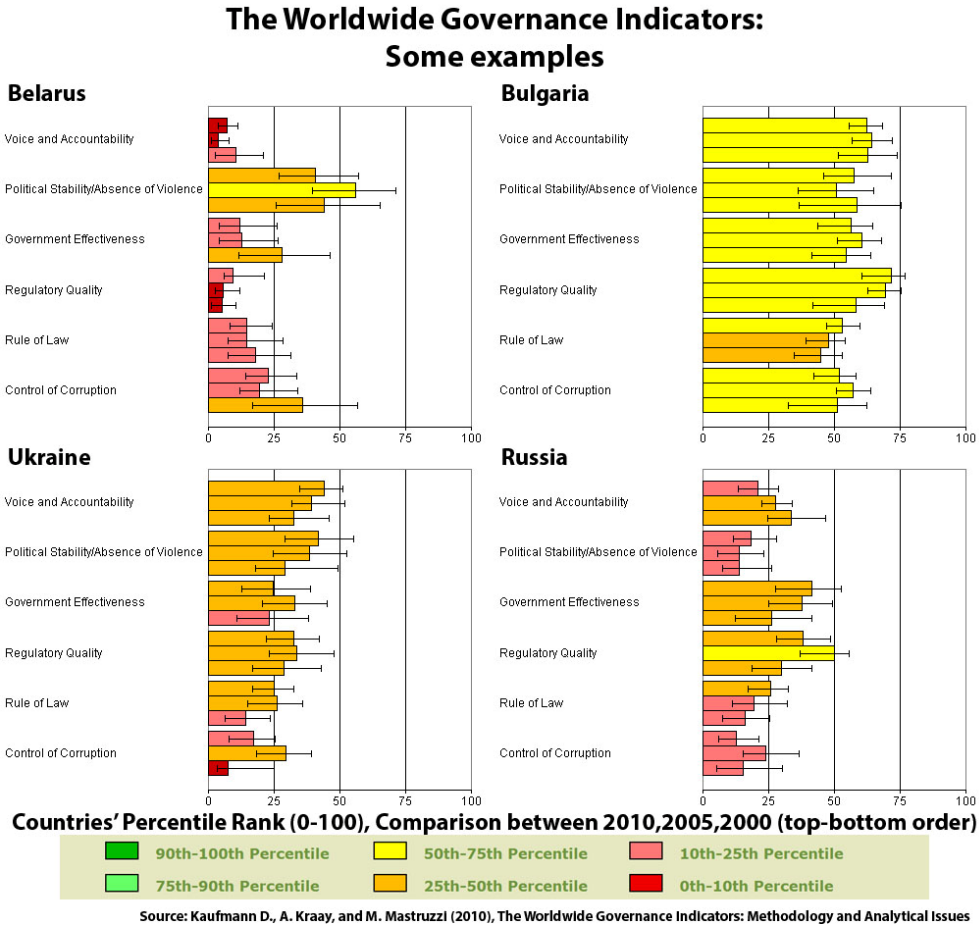

Figure 8. World Bank Benchmarking Work on Quality of Governance

One of the most interesting recent trends—also from NATO's point of view—is the fact that many international organizations (World Bank, International Monetary Fund [IMF], OECD, etc.) have picked up benchmarking as a standard technique to track countries' or regions' progress on various policy issues, even difficult ones such as education, health care, or corruption (see Figure 8). This trend goes back to at least the 1960s when the International Association for the Evaluation of Educational Achievement produced its first international rankings of school mathematics attainment. The World Economic Forum (WEF) has been producing its well-known international rankings of competitiveness since 1979. And over the past two decades many new international rankings have been introduced, including the United Nations Development Program (UNDP) Human Development Index (introduced in 1990), Transparency International's Corruption Perception Index (in 1995), the international health survey produced by the World Health Organization (in 1995), and the OECD Program for International Student Assessment (PISA) rankings (in 2000). As one author noted:

> *[Y]ou can scarcely pick up a newspaper today without reading that your country rates third in this or fifteenth in that, has slipped back or climbed up the rankings for transparency, or competitiveness, or health, or crime, or school attainment, or e-government. Political incumbents use upward movement or high positions in these rankings as opportunities to claim credit while challengers use downward movement or unfavourable rankings to lay blame. News media highlight surprising or dramatic*

ranking outcomes. The policy wonks in strategy units working for government leaders mull over the numbers[36].

This upsurge in systematic data-driven comparative work by international organizations has enhanced both policy transfer and policy learning across countries:[37] "a process in which knowledge about policies, administrative arrangements, institutions, etc. in one time and/or place is used in the development of policies, administrative arrangements and institutions in another time and/or place."[38] In essence, this approach offers an evidence-based alternative to developing new programs or policies as it is based on programs that might have been operating for a long period of time elsewhere—something not typically the case with lessons learned from one's own experiences, let alone "new" initiatives.[39]

One of the most striking examples of this form of benchmarking is probably the work of the OECD, an international organization that regularly publishes benchmark studies on a variety of different policy issues (e.g., in the fields of education and health care). The OECD's website explains its current mission as "promot[ing] policies that will improve the economic and social well-being of people around the world." And it very simply yet elegantly states that one of the ways in which it pursues that mission is by providing "a forum in which governments can work together to share experiences and seek solutions to common problems."[40] This is how the organization describes its own peer review process: "Among the OECD's core strengths is its ability to offer its 30 members a framework to compare experiences and examine "best practices" in a host of areas from economic policy to environmental protection. OECD peer reviews, where each country's policy in a particular area is examined by fellow members on an equal basis, lie at the heart of this process. A country seeking to reduce unemployment, for example, can learn valuable lessons from its peers on what has worked and what has not. This can save time, and costly experimenting, in crafting effective national policies. The recommendations resulting from such a review can also help governments win support at home for difficult measures. And perhaps most importantly, because everyone goes through the same exercise, no country feels it is being singled out. Today's reviewers will be in the hot seat themselves tomorrow."[41]

Much of the analytical work behind this peer review is done by OECD staff. One of the most useful aspects of this work is that it does not just merely provide rankings on various output measures, but also detailed and careful evidence-based comparisons of the various different policy choices (throughput) that countries have made in a number of policy areas. Figure 9, for instance, shows some results on both ***inputs*** into education policy (the horizontal axis indicates the normalized amount of money countries spend on education) and on ***outputs*** (the vertical axis shows countries' students' performance on a standardized science test)[42]. This particular graph shows that certain countries (e.g., the United States and Norway) spend more

---

[36] Hood, Dixon, and Beeston, "Rating the Rankings."

[37] Dolowitz and Marsh, "Who Learns What from Whom"; Dolowitz and Marsh, "Learning from Abroad"; Knill, "Introduction"; Malik and Cunningham, "Transnational Policy Learning in Europe."

[38] Dolowitz and Marsh, "Who Learns What from Whom."

[39] Rose, "Ten Steps in Learning Lessons from Abroad."

[40] "About OECD."

[41] Organisation for Economic Co-operation and Development, *Peer Review*.

[42] The OECD Programme for International Student Assessment http://www.pisa.oecd.org/pages/0,3417,en_32252351_32235731_1_1_1_1_1,00.html.

money on education than most others, and yet score lower on science performance than countries who spend significant less (like Australia, Japan, the Netherlands, and especially Finland).

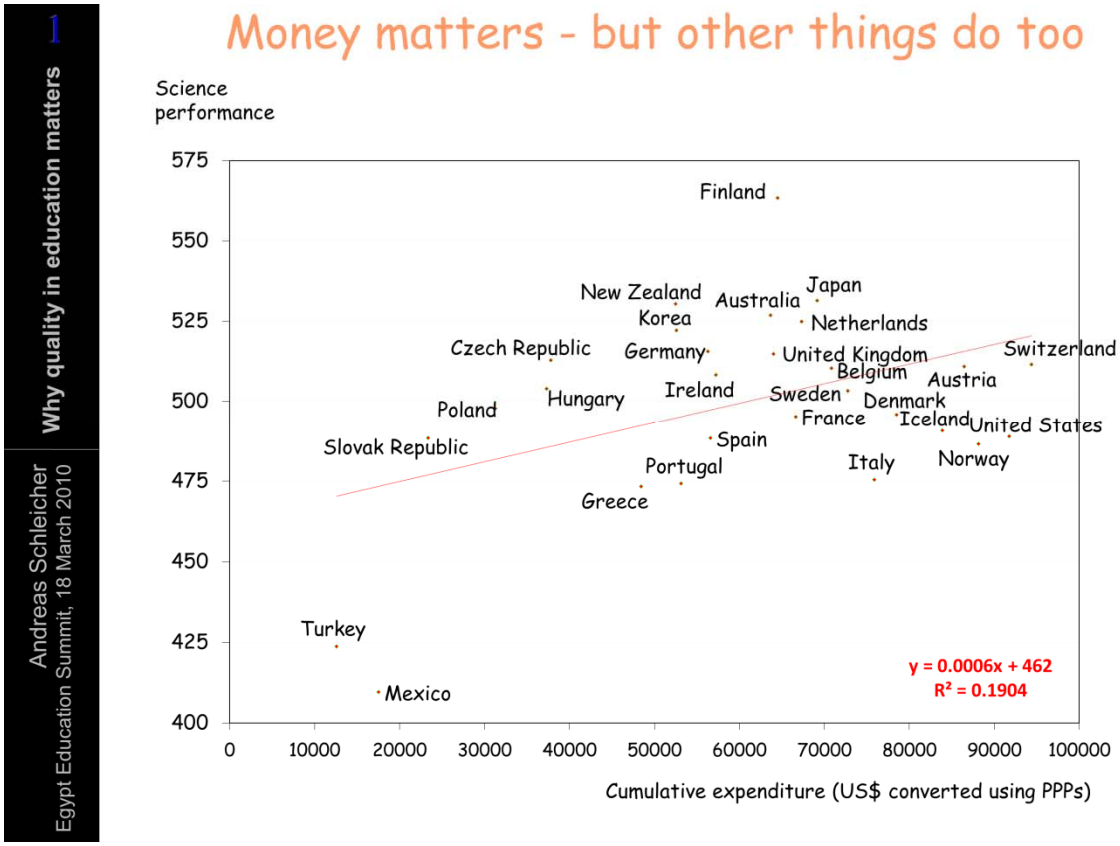

Figure 9. OECD Benchmarking Work on Education

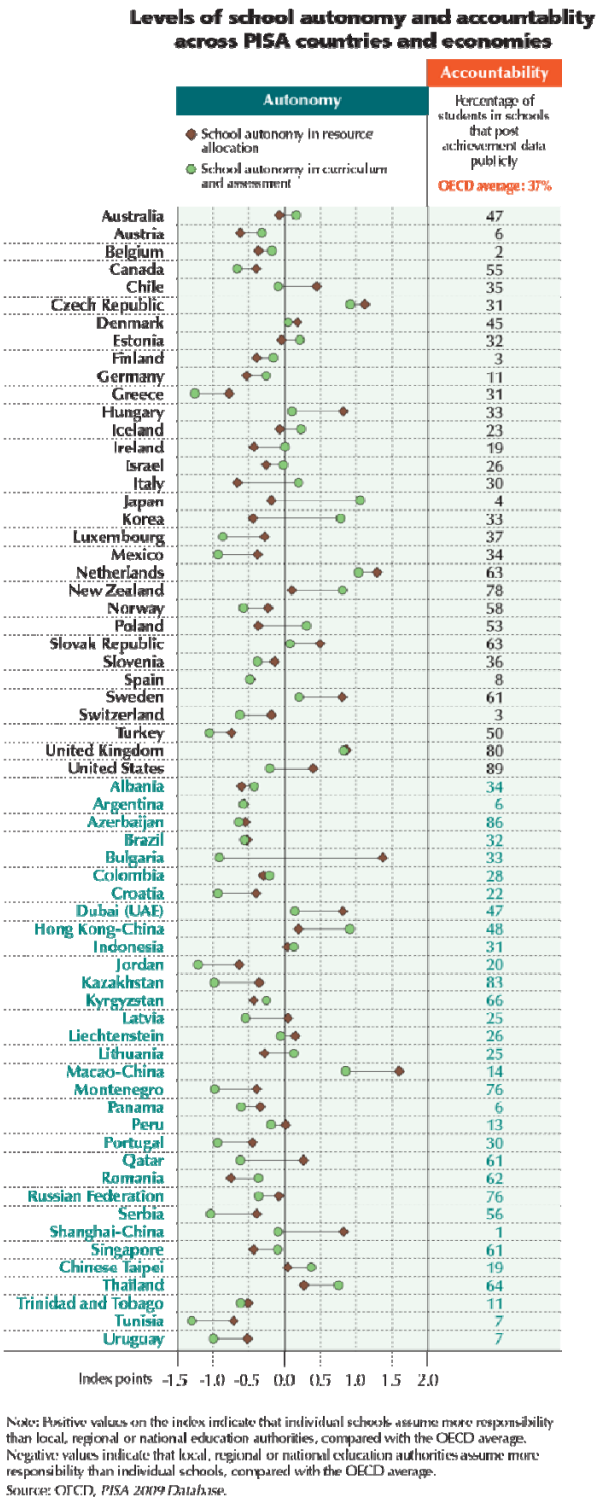

Figure 10. OECD Benchmarking Work on Education

No policymaker or politician (or concerned citizen for that matter) can look at this graph and resist the temptation to identify where his or her country ranks. And invariably this will raise questions like "What does Finland do differently in order to score so unusually well on science despite spending only a comparatively moderate amount of money?" And on this question too, OECD studies provide a number of clues by digging deeper into the various policy choices that have been made by countries in these policy areas. One of the tools the organization uses is TALIS (the OECD Teaching and Learning International Survey[43]). It maps working conditions of teachers and the teaching and learning practices in schools in 24 countries across 4 continents. As an example, Figure 10 shows how much autonomy schools have in the various OECD countries.

It is striking that international organizations like the OECD, EU, World Bank, and IMF are engaging in this type of rigorous evidence-based (and publicly available) analysis for almost all policy areas, except for the area of defense and security.

Evaluating Benchmarking

What have we actually learned from the 25 years of experience we have now accumulated with various forms of benchmarking in the private and the public sector? There is a small but interesting empirical body of literature on the actual practice of benchmarking across different sectors. This section will succinctly present some of the main findings of this literature, based mostly on the most complete recent dataset we were able to identify[44].

[43] See http://www.oecd.org/edu/talis.

[44] Adebanjo, Abbas, and Mann, "An Investigation of the Adoption and Implementation of Benchmarking."

Fields of Benchmarking

Figure 11 shows that benchmarking is being used in many different sectors, with manufacturing still on top, but an increasingly broadening array of other sectors also well represented (including government administration and defense—although the data do not allow us to identify how large the "defense" subset is in this sector)[45].

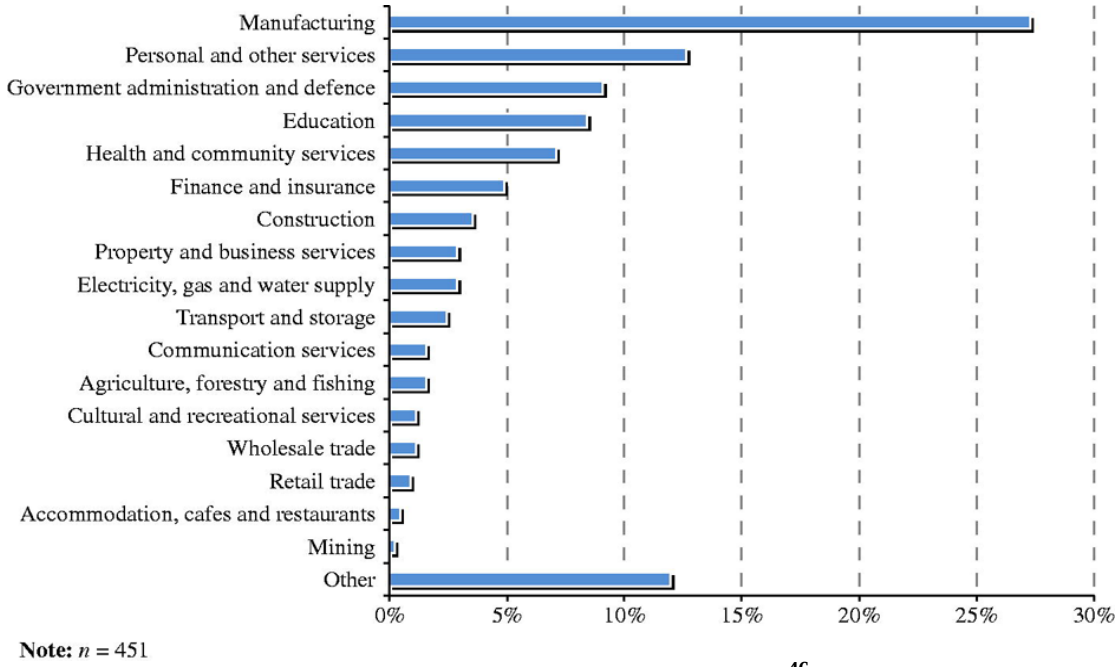

Figure 11. Fields of Benchmarking[46]

[45] The author expresses his gratitude to Dr. Dotun Adebanjo and Dr. Robin Mann from the Centre for Organisational Excellence Research (COER), Massey University, New Zealand for providing him access to the data set they collected.

[46] Adebanjo et al., "Twenty-five Years Later–a Global Survey of the Adoption and Implementation of Benchmarking."

Motives for Benchmarking

Figure 12 illustrates that enhancing one's performance is by far the dominant main driving force behind benchmarking.

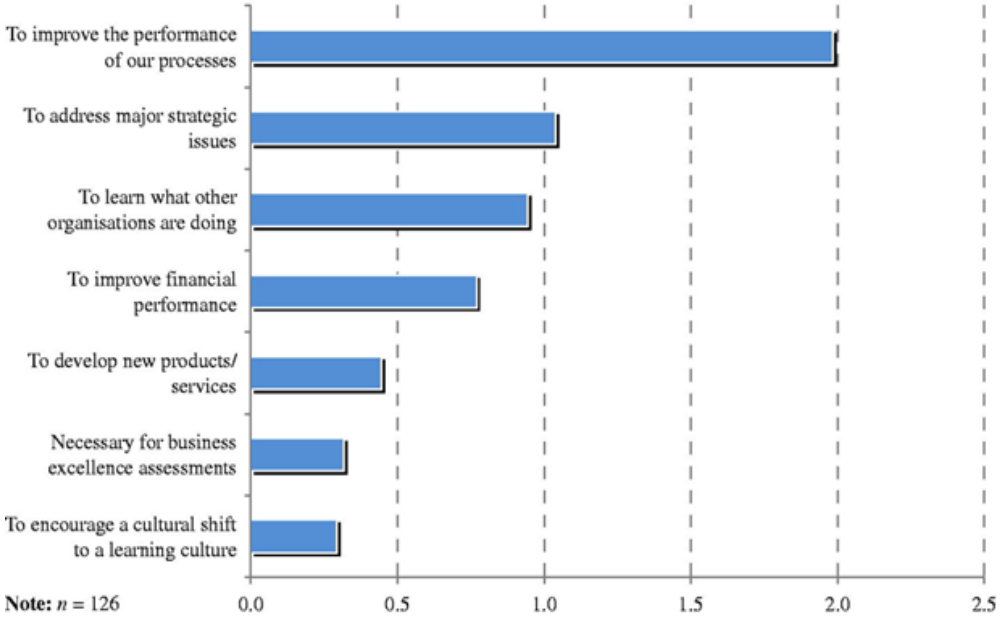

Figure 12. Motives for Benchmarking[47]

Size of the Benchmarking Team

The graph in Figure 13 shows that benchmarking efforts within organizations do not necessarily require large dedicated teams, but can be successfully executed with a small "hard core" that can then be augmented by specialists from throughout the organization for the topics that are being benchmarked.

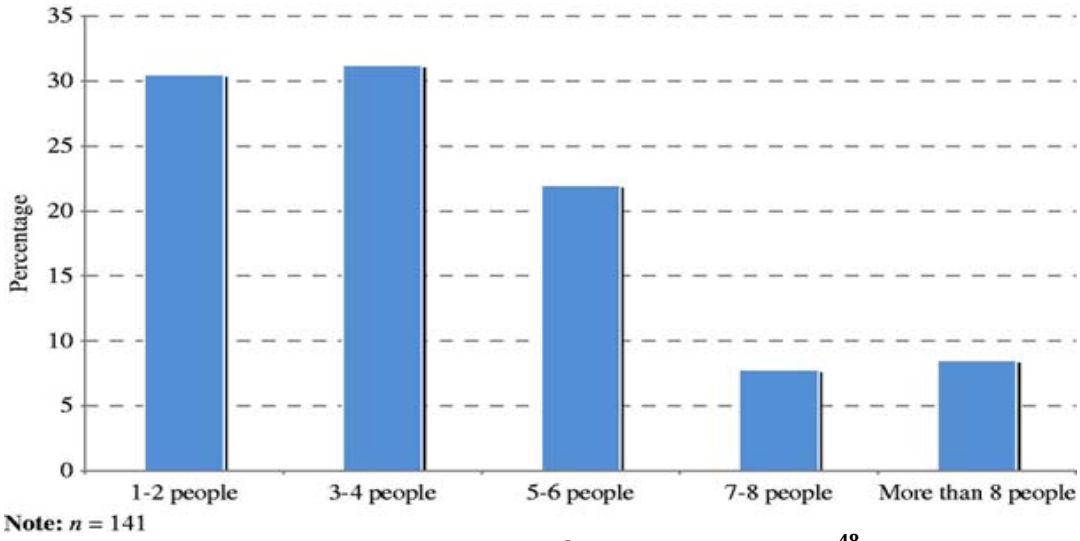

Figure 13. Typical Size of Benchmark Teams[48]

---

[47] Ibid.

[48] Ibid.

Time Required

As with the previous figure, Figure 14 shows that benchmarking projects do take some time, but that two-thirds of all projects in this sample were completed within 4 months.

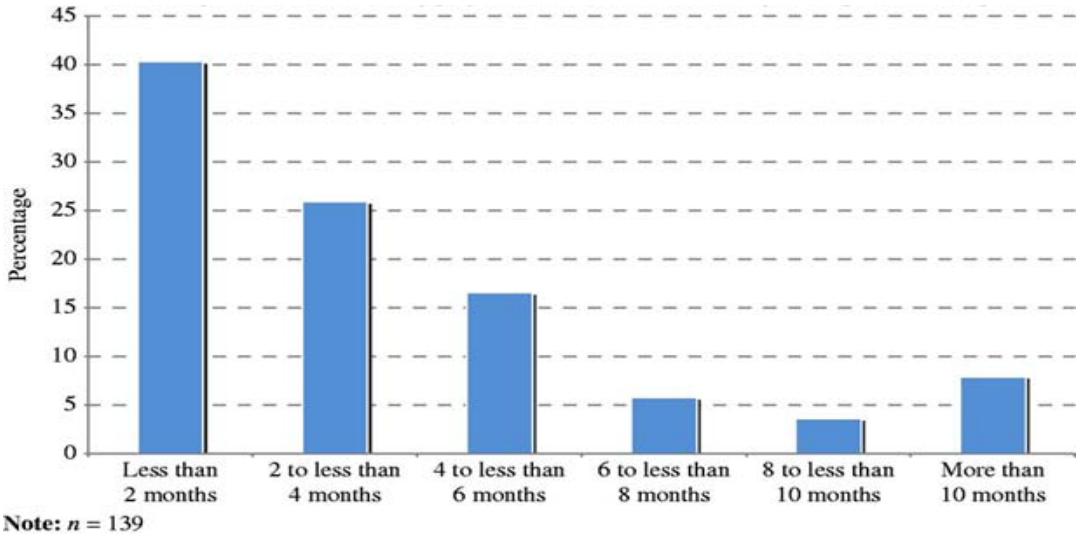

Figure 14. Typical Time for a Benchmark Project[49]

[49] Ibid.

Effectiveness of Benchmarking

The graph in Figure 15 indicates that organizations felt that certain forms of benchmarking were not the most effective techniques for improving organizations. But still about two-thirds of the organizations that participated in this survey claim that their organization's benchmarking projects had proved effective.

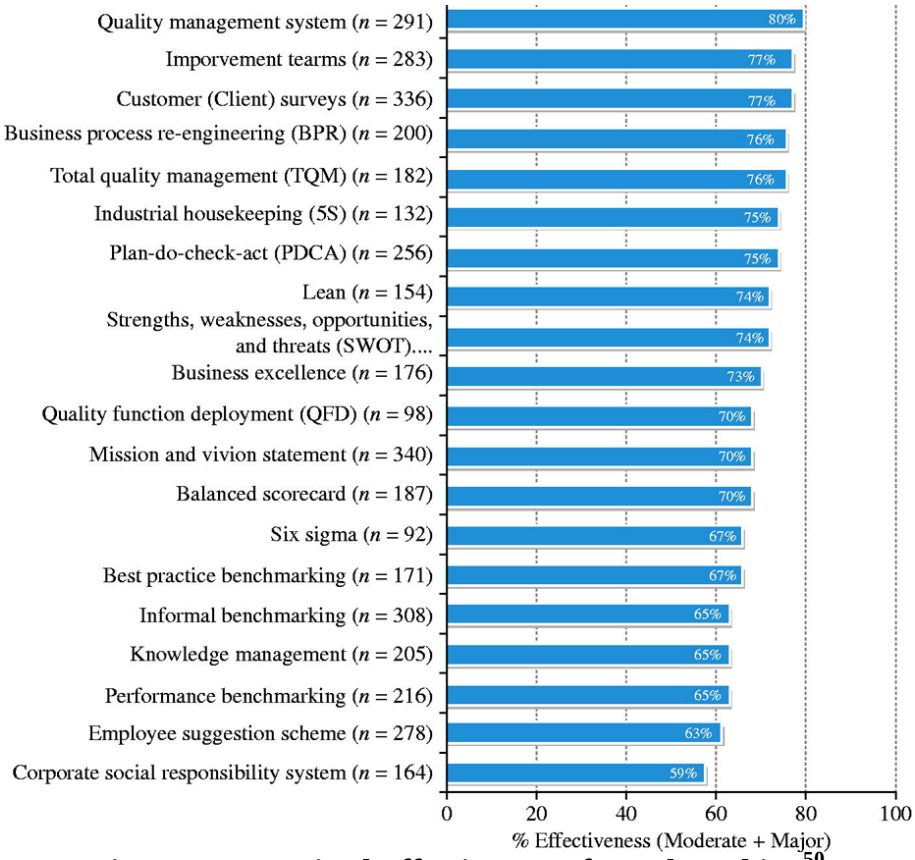

Figure 15. Perceived Effectiveness of Benchmarking[50]

Benefits of Benchmarking

A variety of studies have shown a strong direct link between benchmarking and improved operational and business performance in the private sector.[51] In the public sector, the evidence is less convincing, but this may be attributable to the fact that public benchmarking has not been practiced systematically for quite as long. But here too, the swelling uptake of the technique in the public sector does suggest that many organizations at least anticipate some benefits. A 2001 survey saw the benefits distributed in the way described in Table 1.

---

[50] Ibid.

[51] Voss, Åhlström, and Blackmon, "Benchmarking and Operational Performance"; Ulusoy and Ikiz, "Benchmarking Best Manufacturing Practices;" Sommerville and Robertson, "A Scorecard Approach to Benchmarking for Total Quality Construction;" Adebanjo, Abbas, and Mann, "An Investigation of the Adoption and Implementation of Benchmarking."

Benefit	No answer	None	Some	Moderate	High	Very high
Quality improvement	5	16.3	26.7	30.2	17.8	4
Increased speed of service	5.5	26.9	21.9	26.4	12.9	6.5
Innovate approaches to business improvement	4.5	19	34.5	27	13	2
Process improvement	3.5	6.5	26.4	27.4	28.4	8
Understanding of customer requirements	4.5	15.4	23.4	25.9	23.9	7
Setting of internal standards	5.5	11.9	20.9	33.3	23.4	5
Influencing the strategic decision-making process	3.5	7	25.9	22.4	29.4	11.9
More effective and efficient management of resources	5	15.4	19.4	29.9	20.9	9.5
More effective deployment of resources	4	8.4	18.3	31.7	28.7	8.9
Improvement in people management	4.5	12.4	22.4	31.8	20.9	8
Change in approach style of leadership within the organization	5	14.4	23.8	29.7	22.3	5

Table 1. Benefits of Benchmarking[52]

Another indication of the perceived benefits can be gleaned from the stated intention to use various improvement techniques. Whereas we saw in Figure 15 that benchmarking scored well as a current improvement technique in absolute terms, but scored lower relatively to other techniques, Table 1 shows that when polled about future benchmarking intentions, benchmarking scored better than the other techniques.

Figure 16. Future Use of Improvement Techniques[53]

[52] Jarrar and Zairi, "Future Trends in Benchmarking for Competitive Advantage."

[53] Adebanjo et al., "Twenty-five Years Later–a Global Survey of the Adoption and Implementation of Benchmarking."

We have found no statistical or econometric studies that tried to empirically demonstrate any link between benchmarking and performance. But the stated preference for this technique that comes out of these data combined with the revealed preference of these companies actually continuing to engage in it does suggest that they at least perceive benchmarking as worthwhile.

BENCHMARKING IN DEFENSE ORGANIZATIONS

Just as the public sector lagged behind the private sector in its adoption of benchmarking, so too do defense organizations run behind a number of other public sector domains. This section will therefore first provide a brief state of the discipline of defense benchmarking and will then describe in some more detail what we consider to be two best-of-kind examples of defense benchmarking: a large 2010 benchmarking study by McKinsey and the mainstreaming of defense benchmarking throughout the Netherlands Defense Organization.

Defense Benchmarking – The State of the Discipline

As part of a larger study commissioned by the Dutch Ministry of Defense, TNO (the Dutch Research and Technology Organization) in 2006 identified and analyzed some 200+ publicly available defense benchmarking studies in the area of defense.[54] For this analysis, a template was made for every defense benchmark study containing information about areas such as the background of the study, the "customer," the "executor," the year of publication, the topic, the source (and the actual full text of the study), but also the type of benchmark, the "solidity" (based on some criteria), the cost (rarely available), the timeframe, and the outcome.

To the best of our knowledge, this effort remains the only attempt to take stock of various experiences with defense benchmarking. The study itself is not publicly available, but we will briefly summarize some of the main findings of the analysis.[55]

The analysis showed that while many defense organizations pay frequent lip service to benchmarking, **"real" benchmark studies are few and far between**. The TNO team scanned the Internet for all publicly available documents containing the words "benchmarking," or "benchmark" and "defence," or "defense." That initial search yielded some 1000+ documents that showed some similarity to a benchmarking attempt in the sense defined in this report (i.e., an at least somewhat methodologically conscious attempt at evidence-based comparison of some aspect of the defense organization). A closer look at these documents, however, showed that only about 100 documents actually contained real systematic comparisons.

Of those real benchmarking exercises, the overwhelming majority were **internal** benchmarks (e.g., comparing bases within a country, or processes between a country's services). Less than 5 percent consisted of **external** ones (i.e., where certain aspects of a defense organization were compared with the defense organizations of other countries or with other [non-defense] organizations).

Most of these external exercises tended to be "quick and dirty." In many cases, these external benchmarks were a (small) part of a larger research study on some aspect of a defense organization, where the international (or external) comparison seems to have been not much more than an afterthought. Often the international comparison part of this study consisted of a

---

[54] De Spiegeleire, *Towards a Benchmarking Methodology for Defence*.

[55] The author of this study, also the principal investigator for the 2006 TNO defense benchmarking studies, gratefully acknowledges the willingness of the Dutch defense organization to share this work more broadly.

few contacts with others or some input by local defense attachés from other countries about the issue at hand, yielding brief parallel descriptions of others' experiences with the topic at hand, but without a genuine attempt to develop truly comparable metrics. Exceptions to this rule are external benchmarks on processes that are similar to those in the private sector (e.g., logistics), where frequently the expertise of private consultancies with experience in similar processes in the private sector could be drawn upon.

A brighter point was that the study clearly identified an **upward trend** in the number of benchmark studies, reflecting a growing desire by a number of defense organizations to inform their decisions by more systematic comparisons with other countries (or organizations). Although the study only went to 2006, our own anecdotal observations suggest that this upward trend has continued and even strengthened.

Virtually all studies (again with the exception of those that are close to the business world) show **enormous comparability problems**. Although some exercises made attempts to circumvent these, the actual findings of the reports still leave readers with a feeling that the conclusions are only of limited use. Even studies involving relatively easily comparable topics to be benchmarked such as money (see the Danish-Norwegian study on costing, or Stockholm International Peace Research Institute data about military expenditures) had to make enormous efforts to develop genuinely comparable datasets.

Another remarkable observation was that there seemed to be an inverse relationship between the topics that actually are benchmarked and those that probably should be. **Virtually all** external benchmarking exercises tend to be **based on inputs** (e.g., money, people, and systems); far fewer on process **(throughput)**; and **virtually none on outputs** (e.g., operational efficiency) **let alone outcomes/effects.**[56] A trend away from inputs to outputs is discernible, but remains weak.

Finally it was striking that **extremely little information** was available on the resources that had been allocated for the various exercises or on the actual take-up of the studies' findings. The TNO research team even made follow-on calls to many of the organizations or individuals responsible for those defense benchmarking exercises, but even that was insufficient to yield precise figures. All indications are, however, that defense benchmarking (with the possible exception of extremely quick and dirty ones) is currently **quite labor-intensive**, which is not surprising since most of the studies are done in a "unilateral," non-cooperative mode.

The 2010 McKinsey Study

In 2010, the well-known global management consulting firm McKinsey published some information on the world's first large-scale defense benchmark study it conducted between 2008–2009 comparing 33 countries that together account for more than 90 percent of global defense spending.[57] Although not a core focus of McKinsey's activities, the consultancy is

[56] As the defense world starts moving towards effects-based approaches to operations, the pressure for benchmarking to start moving more to the right of this sequence is expected to grow.

[57] "McKinsey on Government. Special Issue: Defence."

still active in the defense and security field in 16 countries with more than 170 engagements between 2006 and 2010.[58]

The publication *Lessons from around the world: Benchmarking performance in defense* contends that performance can indeed be compared across defense ministries wherever they engage in the same types of activities. It presents the important claim that "countries can shrink their defense budgets without losing capability": "Our firm belief is that certain aspects of operational performance are indeed comparable across ministries of defense, and that ministries can learn from one another when it comes to delivering more defense output for the same or less input."[59] What the McKinsey team essentially appears to have done in the study (and unfortunately many details of the methodology have not been made public) is a three-step approach.

First, they collected publicly available hard data on the quantity and type of military equipment, number and general classification of personnel, and annual defense budgets. They disaggregated these data into key spending categories and apparently made an effort to make these data truly comparable (to account for different accounting methods, different size, etc.).

Secondly, they created a new metric for measuring the performance of military equipment, which they called "military equipment output" (MEQ). The metric is a function of four different factors: volume, mix of equipment, age of equipment, and overall equipment quality. Here they appear to have gone to great lengths to make the actual "fighting power" of one military organization comparable to the fighting power of another. In their own description, "[t]he analysis involved using conjoint techniques to assess 69 categories of military equipment across ten countries and five time periods dating back to 1971, generating like-for-like comparisons of the equipment's fitness for purpose. This work produced expert ratings on the overall quality of 5,500 pieces of military equipment—a statistical robustness that gives MEQ much greater reliability than any other published measure of defense output to date."

Finally, they constructed a set of ratios that measure outputs in three core budget areas of defense: personnel, equipment procurement, and maintenance. Table 2 presents data they published with such ratios in those three categories.

[58] Introductory Meeting HCSS – McKinsey & Company, April 2011.

[59] "Lessons from Around the World: Benchmarking Performance in Defense."

Budget area (average % of defense budget)	Key ratios	Range	Average
1 Personnel (45%)	• "Tooth to tail" (combat personnel as % of total personnel)	16–54%	26%
	• Number of deployed as % of total active troops	1–18%	5.3%
	• Personnel costs per active and other personnel	$800–$146,000	$44,800
	• Personnel costs over military equipment output[1]	$2,000–$218,000	$72,000
2 Equipment procurement (18%)	• Military equipment output[1] over procurement and R&D spending (index)	17–330	100
	• Procurement spending over active troops	$1,000–$536,000	$60,000
3 Maintenance (8%)	• Cost of maintenance per unit of military equipment output[1]	$2,000–$104,000	$13,000
	• Cost of maintenance over cost of equipment procurement	8.2–446%	13%

Table 2. McKinsey Defense Benchmark: Ratios in 3 Budget Categories

The quite staggering spreads reported here are impressive indeed. They clearly illustrate that there are enormous differences across countries on some of the most fundamental aspects of defense that deserve to be examined more carefully—along the same lines that pushed Xerox CEO David Kearns to start benchmarking with his Japanese counterparts or that triggered the OECD benchmarks for education or innovation policy. (But contrary to the OECD, McKinsey did not publish any more detailed analyses of these data. This is presumably something it uses in its own engagements with the Ministries of Defense in the countries in which it works).

Another interesting innovation is that for comparison purposes, McKinsey categorized all countries in five clusters based on types of military strategies: global-force projection (countries with worldwide striking capability), small-force projection (NATO members or countries with a fairly significant presence in international missions), relevant national security threat (countries under attack or threat), emerging regional powers, and non-aligned or neutral countries. This allows countries to not compare themselves to all other countries, but also only to their own "peer group."

The main claim of the study is that there remains much scope for streamlining various non-operational activities of defense organizations—essentially by doing similar things to what consultancies have been doing in the private (and increasingly also the public) sector across the world.[60] They cite the example of the defense ministry of "a Northern European nation" that had set itself a goal to "increase its tooth-to-tail ratio from 40:60 to 60:40 over three years. It achieved this goal by centralizing formerly duplicative support functions including Human Resources, Information Technology, finance, media and communications, health services, and facilities management. By mapping the functions' activities and resources—what exactly each function did, who did it, and how many people did it in each regiment—and by comparing itself with other public and private-sector organizations, the defense ministry realized that centralization would yield savings of approximately 30 percent per function."[61]

[60] "Big Savings from Little Things: Non-Equipment Procurement"; "Mastering Military Maintenance."

[61] "Lessons from Around the World: Benchmarking Performance in Defense," 9.

There are many points in this study that can be criticized. The fact that the data themselves as well as many details of the methodology (e.g., the authors acknowledge that "assembling inputs presented a significant research challenge due to wide variability in the quality and quantity of available data"[62]) were not made publicly available greatly diminishes the effort's authority—despite the impeccable credentials of the organization that stands behind the study. But we do see this study as an impressive first step in the direction of more systematic (and hopefully more transparent) work that remains to be done. More than anything else, this study demonstrates how much can be done even with publicly available data and what types of results such an exercise can yield.

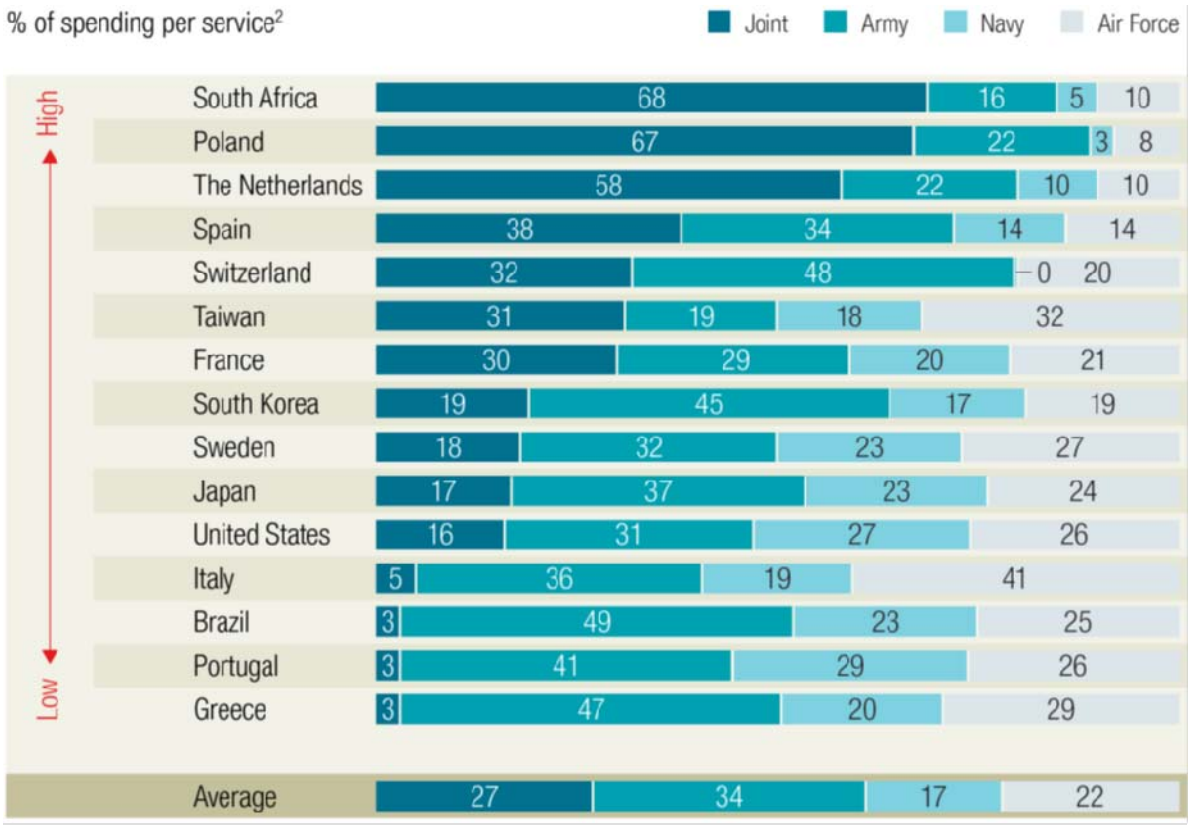

Figure 17. Joint vs. Service Spending

Figure 17 displays the percentage of military spending devoted to joint versus a single military service. In Figure 18, the relative spending levels of each nation's military spending is broken down into combat, combat support, and other. These benchmarks suggest alternative investment options for each country or opportunities to increase their tooth-to-tail ratios and generate more capability from current spending levels.

[62] Ibid., 5.

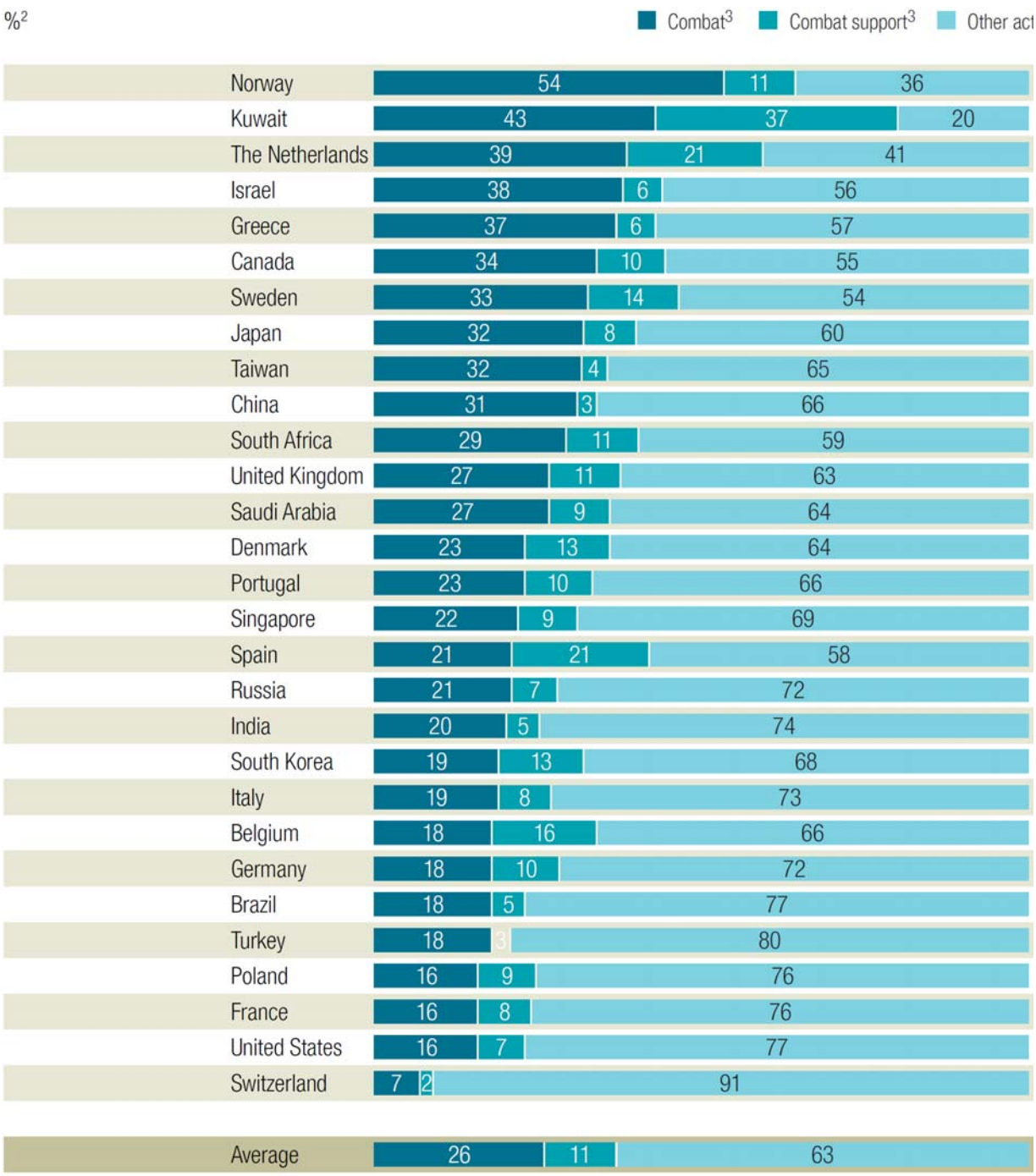

Figure 18. Tooth-to-Tail Ratio

Another useful benchmark involves national force deployability to meet national or NATO requirements for missions beyond national borders. Table 3 displays total and relative levels of deployed and deployable force levels. The last column also reflects the relative costs for troops that are deployed. The data suggest wide variances that could be explored in order to find out what a country like Norway does differently in order to achieve such high deployability in its armed forces.

	Total active (number of people)	Total deployable[1] (number of people)	Deployed (number of people)	Deployed over total active (%)	Deployed over deployable (%)	Cost per troop deployed ($ thousands)
United States	1,352,494	N/A	250,000	18.5	N/A	N/A
United Kingdom	185,950	74,750	34,000	18.3	45.5	N/A
The Netherlands	44,636	17,724	3,896	8.7	22.0	68
Finland	10,100	6,000	840	8.3	14.0	216
Sweden	11,574	3,122	950	8.2	30.4	611
France	262,592	42,500	17,485	6.7	41.1	35
Italy	191,152	54,800	11,170	5.8	20.4	N/A
Spain	77,800	39,617	3,344	4.3	8.4	195
Germany	221,185	37,275	8,946	4.0	24.0	172
Greece	135,500	22,182	1,290	1.0	5.8	83

Table 3. Active vs. Deployable vs. Deployed Troops

The Dutch Approach to Defense Benchmarking

The Netherlands—to the best of our knowledge—is the only country within NATO to have adopted and mainstreamed defense benchmarking throughout the organization and to apply it to all new policy initiatives contemplated by the Netherlands Defense Organization. We will therefore describe the Dutch experience in more detail as a country case study of how one country managed to put this issue on its agenda, studied it, made the decision to embark on systematic benchmarking, and then mainstreamed it throughout the organization. We will also provide a concrete example of a larger Dutch benchmarking study on the topic that lies at the heart of this paper: how a number of defense organizations (and one non-defense organization) translate policy ambitions into capabilities.

Background

The Netherlands' Ministry of Defense (MoD)—as most of its peers—has always shown a keen interest in learning from the best. It should thus not come as a surprise that the organization has over the years engaged in various forms of defense benchmarking even if those efforts were not always given that name. Around mid-2004 the issue of benchmarking started gathering new momentum within the defense organization.

MoD's Policy Planning Staff (the Directorate of General Policy Affairs [in Dutch HDAB]) decided to perform an inquiry into the ways in which defense benchmarking was being done within the Ministry. It came to the conclusion that there was no standard or broadly applicable benchmarking method[63] and that it might be worthwhile to investigate whether such a method was feasible and desirable. In a 2004 note, HDAB spelled out its thinking:

> *The Dutch Armed Forces are internationally oriented and embedded – operationally, managerially and organizationally. Within the context of homeland security tasks, the organization is increasingly intertwined with other departments, other levels of government and non-governmental actors. In many fields, the defense organization furthermore increasingly interacts with the private sector for various materiel and personnel issues. Policymakers within the Defense organization increasingly have to take these developments into consideration. This includes keeping track of relevant*

[63] Based on a poll by HDAB in 2005 with DS/DOBBP/OB/TV, DGFC/DBE and DMO/DR&D.

developments, knowledge, expertise and experiences within the aforementioned – and possible even other – partners. The aim is among other things to acquire better insights into the operational effectiveness of the Armed Forces and to identify potential inputs into the Policy, Planning and Budgeting process. It can be used for widening the possibilities to come to an exchange of best practices internationally, interdepartmentally, or in civil-military terms. Furthermore it can contribute to the development of target metrics for the deployability of the Armed Forces. These metrics are important nationally, as in the monthly reporting by the department, and internationally as in the development of usability criteria for NATO forces.

This note was discussed within the Department's Policy Council (the highest-level policymaking body), which agreed to embark upon a serious benchmark study comparing the Dutch Armed Forces with other Armed Forces. In first instance, the focus of the benchmark was intended to be the operational effectiveness of the Armed Forces. Benchmarking was seen as an instrument that could assist in improving the Armed Forces' effectiveness and efficiency. The discussion within both NATO and EU about *output*- and *usability* criteria clearly played an important role in this. The envisioned benchmark study was included in the MoD's Policy Vision 2007 as a matter for further policy development in 2005–2006. HDAB presented its vision on how to proceed with this initiative before the Policy Coordination Council, which approved the plan and recommended swift implementation. It also suggested that the scope of the method be expanded to include comparability with non-defense organizations.

Then in 2006, MoD commissioned the Dutch national Research and Technology Organization TNO, housing about 1,000 defense and security scientists, to conduct a study examining the feasibility of developing a generic defense benchmarking method. To this end, a Benchmarking Working Group was created with representatives of the five main components of the Dutch Defense Organization: the Directorate of General Policy Affairs (HDAB), the Chief of the Defense Staff (CDS), the Directorate-General of Finance and Control (DGFC), the Directorate of the Defense Materiel Organization (DMO), and the Directorate of Personnel (HDP), which also includes the Directorate for Healthcare (DMG).[64]

TNO Report on Defense Benchmarking: A Double Recommendation

The TNO report was delivered in late 2006. It contained the analysis of the State of the Discipline in Defense Benchmarking that was already referenced in the previous section. The report concluded with a double recommendation to the Dutch MoD.

Given the difficulties surrounding defense metrics in general, and specifically *comparable* defense metrics, the **first and primary (more long-term) recommendation** was to work towards a convergence of defense performance management practices—in essence a cooperative and multilateral approach to the issue:

A genuine and reliable benchmarking methodology can in our view only emerge from a comprehensive attempt to synchronize various trends in many defense establishments inside and outside of NATO towards 'modern' internal performance evaluation and management. To date, these trends remain purely national. Even those countries that are adopting a similar methodology (family) for this internal

[64] The author of this study was the lead for the study on the TNO side.

performance management (such as the various versions of the 'Balanced Scorecard'-methodology) still use widely differing categories, performance indicators and metrics. Because of the significant difficulties in introducing these new management systems (given the multiple defense information systems that tend to exist in most countries), external comparability with other defense organizations tends not to be a consideration. It stands to reason that the transition to national unified defense information systems provides a unique window of opportunity to also synchronize these multinationally. In many cases some reflection on the issue of external comparability might even yield a better internal performance indicator.

The study concretely pointed to three ongoing international efforts to work towards such synchronization:

- The "***Community of Practice on Defense Performance Management***," an informal framework that was that was initiated by the Canadian MoD in October 2004 (based on the Technical Cooperation Programme (TTCP[65]) countries and a few selected other national defense organizations [NDO]) and was picked up by the British MoD in December 2005. As a result of this TNO recommendation, the Netherlands became an observer nation in 2005 and has been a full-fledged one since 2006, organizing the meeting itself in 2007 around the very topic of benchmarking.
- Danish-Norwegian efforts (*Denmark-Norway Comparative Study*[66]) to develop a model for the comparative analysis of the defense sector in those two countries, focusing primarily on the comparability of the available financial data.
- The NATO Research and Technology Organization's [67] proposed System Analysis and Studies (SAS) panel on costing that was being stood up to estimate and compare defense costs. This effort would become SAS-076 *NATO Independent Cost Estimating and its Role in Capability Portfolio Analysis,*[68] in which the Netherlands (again on the basis of the recommendation contained in the TNO report) became an actively participating member.

The authors of the report were under no illusion that any of these more cooperative and multilateral efforts, however worthwhile in their own right, would lead to any great breakthrough in the near- to mid-term. Based on the critical, but on balance, still positive analysis of the promise of defense benchmarking, the team therefore also developed a generic defense benchmarking planning guide that was intended to enable meaningful unilateral comparisons even in the absence of genuinely comparable data sets.[69] The second shorter-term policy recommendation was therefore to explore whether the planning guide could be turned into a more permanent defense benchmarking method. The report recommended identifying a number of issues for pilot benchmark studies and to then reconvene the Benchmarking Working Group to decide on further steps.

---

[65] "The Technical Cooperation Program (TTCP)."

[66] Berg-Knutsen and Østbye, "Economic Analysis at FFI."

[67] "NATO Research & Technology Organisation."

[68] Available at http://www.rta.nato.int/Activity_Meta.asp?Act=SAS-076.

[69] 'Unilateral' in the sense that the Netherlands would proceed with the benchmark study even in the absence of any active cooperation of the other organizations against whom the Netherlands would be benchmarked.

TNO Defense Benchmarking Planning Guide

Learning from a number of both good and bad practices in the world of public (and private) benchmarking, the TNO method prescribed a number of steps, tips, and tricks that were intended to help defense organizations in teasing out interesting and useful lessons from other referents.

The detailed description of the actual method (including how it was developed) is contained in two more detailed (but non-public) TNO reports that were written in English: *Towards a Benchmarking Methodology for Defense* (2006)[70] and *Learning to Learn Validating the TNO Defence Benchmark Planning Guide*[71]. For the purposes of the current study, we will present some of the main defining features of the TNO approach:

- **Systematic "*topic-to-metric*" decomposition (also for "soft" issues)**: The method emphasizes that benchmarking requires metrics—common yardsticks along which the differences between referents can be presented in a clear (both logically and visually) way. It contains a number of tips and tricks on how any topic can be decomposed in a number of categories for which one can identify indicators that can be expressed in metrics—sometimes hard, sometimes soft. Figure 19 provides an example from a benchmark of national security strategies (NSS), which were decomposed in a number of categories that were found back in most NSS, and then further decomposed into concrete indicators that were operationalized to some metrics (in this case, for instance, the semantic weight throughout the individual NSS of certain baskets of words, such as those related to "military" tools, as determined by a text mining tool).

Figure 19. Example of the Topic-to-Metric Decomposition Approach

- **Structured method (*step-by-step planning guide*):** Based on an analysis of more than 200 defense benchmarking exercises, the method spells out a protocol with a number of sequential systematic steps that can help in coming to useful findings. An important part in this protocol is that it starts with a smaller feasibility study based on a quick scan of the available information leading to a go/no go decision point.
- **Based on *primary sources* (*not* phone calls, questionnaires, or benchmarking tourism)**: The method strongly favors using authoritative documents as a basis for the

[70] De Spiegeleire, *Towards a Benchmarking Methodology for Defence*.

[71] De Spiegeleire and Jadoul, *Learning to Learn Validating the TNO Defence Benchmark Planning Guide*.

benchmark study (especially since MoDs typically codify and document many of their activities) over more subjective information (however potentially insightful).

- **More about mapping differences than about a beauty contest (*descriptive, not normative*):** Given the current sorry state of standardized metrics in defense (especially on outputs), it is often impossible to make value judgments about different choices made by referents. But the method strongly argues that even just mapping differences between referents can prove extremely instructive (More on that later in this paper).
- **Strong recommendation to include at least one *non-military referent***: Avoid the temptation to claim that "defense is totally different" (and as a corollary "can therefore not be compared with non-military referents"). The method argues that the benefits of considering outside organizations or businesses and analyzing these along the same lines as defense outweigh the drawbacks (especially when the protocol for selecting referents is applied judiciously).
- ***Spiral development* instead of *rigorous sequentialism***. Given the many uncertainties that often accompany the quest for information about the referents, the method advocates adaptiveness throughout the process.

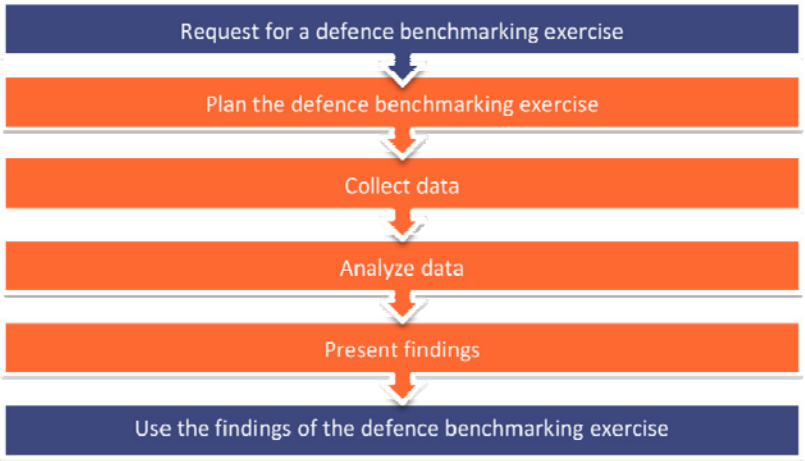

Figure 20. The Main Stages of the TNO Defense Benchmarking Planning Guide

Mainstreaming the Method

In 2007 and 2008, six pilot benchmark studies were undertaken by the Dutch MoD and TNO. They ranged from some quite concrete studies ("Forward Tactical Medevac" and "Large Complex Critical Infrastructures" to broad ones such as "Effects-Based Approached to Operations," "Network-Centrism," "National Security Strategies," and "Output Steering").[72]

As a result of these pilots and the report TNO produced on them[73], in 2008 the highest policymaking-body in MoD decided to consider the *TNO Defense Benchmarking Planning Guide* as validated. The highest civil servant in MoD, the Secretary-General, mandated a benchmarking study (at least a benchmarking feasibility study) for all major policy decisions made by the defense organization. The *TNO Defense Benchmarking Planning Guide* was

[72] A benchmark study for the EU 6th Framework Programme Research IRRIIS project – Integrated Risk Reduction of Information-based Infrastructure Systems. "IRRIIS – Integrated Risk Reduction of Information-based Infrastructure Systems."

[73] De Spiegeleire and Jadoul, *Learning to Learn Validating the TNO Defence Benchmark Planning Guide*.

made available throughout the department and was also complemented with a defense benchmarking "wiki". The department furthermore instituted a biannual "Quality of Policy" training course for (each time) about 15 to 20 MoD staff members (both military and civilian and from throughout the organization) in which an entire half-day is devoted to instruction on defense benchmarking. Many parts of the organization have completed real defense benchmark studies since then, from fairly modest ones to sizeable ones.

One of the most influential uses of benchmarking came during the Netherlands' big bottom-up defense review, which contained a number of interesting benchmarking data[74] and also drew upon the insights derived from the larger defense planning benchmarking effort that will be reported upon in the next section (Capability Development) of this paper.

Today, about 100 people within the Dutch defense organizations have had first-hand experience with the *TNO Defense Benchmarking Planning Guide*. Many lessons, both positive and negative, have been learned. The two main critical issues that we would like to flag in this paper are:

- Difficulties in **collecting the data**—The planning guide is in essence for unilateral benchmarking, which makes it much harder to ensure access to the written (and thus officially approved) documents required for the systematic analysis that is advocated;
- **Commitment** from the participants who have to do the work—As we also saw in the evaluations of various non-defense benchmarks, completing a meaningful benchmarking study is labor-intensive and far from trivial. Here too, the fact that the Dutch method is unilateral poses various challenges that could more easily be overcome in a more multilateral setting.

We still take comfort in the thought that despite these difficulties "benchmarking in one country" continues to enjoy broad support throughout the defense organization. The way the planning guide is structured now, a (mandatory) small preliminary feasibility study has to be executed for every policy initiative to see whether the anticipated benefits of a more rigorous benchmarking study exceed the anticipated costs of a more thorough study. We see the fact that some groups do proceed with a "full' benchmark study as proof that even unilateral benchmarking can be made to work, which bodes well for more cooperative forms of defense benchmarking. But even the feasibility study itself is seen by many as a useful impetus to also look outside of the organization for inspiration at the outset of a new policy decision.

Dutch Example of a Defense Benchmarking Study: Capability Development

In 2007–2008, HCSS was commissioned by the Dutch MoD to benchmark the ways in which a number of countries derive their military capabilities—the topic that lies at the heart of this paper's call to take the battle upstream. We will report here on some findings of that study in order to provide concrete illustrations of:

- the ***types of data*** that can be used and /or collected (at minimal effort or cost)—***input***;
- the ***procedures*** that can be used to make data comparable—***throughput***; and the
- ***types of results*** that can be expected from benchmarking—***output***.

[74] Ministerie van Defensie, *Verkenningen Houvast Voor De Krijgsmacht Van De Toekomst*.

As we illustrated in Figure 4, the capability development process remains an essentially national one on which (for most member countries) NATO's impact is fairly minimal. In its most general form, this process is fairly similar in all countries and consists of two main steps:

- First, a country's highest national political leadership defines what it wants to use its armed forces for (ambition) and specifies the budgetary envelope within which this ambition has to be realized (**high level policy parameters**).
- Then, the NDO takes this high-level political guidance and converts it into a number of **concrete capability choices.**

The precise ways in which these two general steps are implemented vary quite significantly across NDOs. It is fair to say that most countries struggle with the translation of (typically fairly abstract) policy guidance into concrete capabilities. Larger NDOs tend to have sizeable staffs (and often analytical support mechanisms and tools from their defense research institutes) to assist them with this Herculean task. Smaller countries tend to have far more modest staffs and support mechanisms. This means that the key decisionmakers in this process have to adjudicate the various pressures coming from numerous powerful parochial interests from the various silos within the NDO, from politics (financial allocation battles, social considerations, regional distribution, ideologies, industrial lobbies, etc.) without analytical counterweights.

For the purposes of this paper, we will just select a few key elements in this process and will analyze how a few countries tackle them. The description will draw heavily from a larger study HCSS completed for the Dutch MoD in preparation for the large bottom-up defense review that took place in 2009–2010[75]. This study was entirely based on publicly available documents: various policy papers (white papers, strategies, etc.), capability development manuals, performance management reports to parliaments and/or accounting chambers. The main purpose of the study was to present the Netherlands Defense Organization with a number of findings from other countries or organizations that could be processed into a new, more systemic approach to strategic defense management integrating strategic (political) choices, resource allocation, capability planning, and performance management. In search of such "nuggets," HCSS worked in close cooperation with some key NDO players and studied a number of countries (Australia [AUS], Belgium [B], Denmark [DK], France [F], and the United Kingdom [UK]) and one international organization the World Food Programme [WFP], an operational organization that is also engaged in the very same crisis zones as defense organizations). They benchmarked the ways in which these countries and organizations a) set their defense ambitions, b) translate those (often abstract) ambitions into real-life defense capabilities, and c) then managed the performance of the resulting armed forces. This chapter also benefited from the author's participation in an ongoing benchmarking effort of capabilities-based planning within the Technical Cooperation Program—the "five eyes" equivalent of NATO's Research and Technology Organization.

Level of Ambition

We already pointed out political guidance plays a central role in providing the high-level policy parameters for real defense capabilities. At first glance, one might ask how such an

[75] De Spiegeleire et al., *Closing the Loop. Towards Strategic Defence Management*.

abstract, "political" element could possibly be benchmarked. Yet that is precisely what the HCSS benchmarking team set out to do. In this report, we will focus on two aspects of this ambition level: its substantive content and its level (ambitiousness).

The Content of Ambition

The HCSS team analyzed patterns and trends in the ways in which the ambition level is described in the high-level documents. This was done on the basis of the following four categories:

1. **What:** Comprised of parameters that specify important elements at the core of defense policy such as interests, principles, vision, various threats that have to be warded off, and actions that have to be undertaken.
2. **Who:** Consisted of indicators that illustrate the nature of the relationship a referent wishes to have with other nations. These relationships are categorized as unilateral, bilateral, multilateral, and humanitarian.
3. **Where:** Included geographical locations such as regions and countries where referents want to materialize their "What" ambitions. These include national and international.
4. **When:** Focused on indicators that contain a time element such as short or long-term planning horizons. These include time focus and action.

Each category is in turn subdivided into individual concepts and then scored on the basis of a consistent (and transparent) coding scheme. Table 4 presents the findings of our coding of the high-level policy documents around these four categories. To illustrate, within the "what" category, all referents (with the exception of France) claim the ambition of wanting to make the world more secure, whereas the ambition to maintain the free flow of natural resources only really emerged in the second half of this decade.

Table 4. Benchmarking Ambition Levels in Defense White Papers

Ambition	AUS (2000)	AUS (2003)	AUS (2005)	AUS (2007)	B (2003)	B (2008)	DK (2004)	F (2003)	F (2008)	NL (2000)	NL (2003)	NL (2005)	NL (2007)	UK (1998)	UK (2003)	UK (2008)	WFP (2004)
What																	
Interests																	
National interests		X	X	X	X		X	X	X	X				X	X	X	
Economic development														X	X		
Secure world	X		X	X	X	X	X			X	X	X		X	X	X	
Flow of natural resources				X					X						X		
Principles																	
Society				X			X						X			X	
Responsibility	X						X							X		X	
Transparency					X	X											
Human rights						X	X			X				X			X
International law					X	X				X	X	X				X	X
Freedom							X		X	X				X	X		
Protection of allies					X	X					X	X	X				

Ambition	AUS (2000)	AUS (2003)	AUS (2005)	AUS (2007)	B (2003)	B (2008)	DK (2004)	F (2003)	F (2008)	NL (2000)	NL (2003)	NL (2005)	NL (2007)	UK (1998)	UK (2003)	UK (2008)	WFP (2004)
Democracy					X	X	X							X			
Vision																	
Prosperity															X	X	
Leadership				X		X		X				X	X	X	X		X
Force for good														X	X		
Protection																	
Threats (direct/indirect)		X	X	X		X	X	X	X					X		X	
Coercion									X								
Attack	X	X		X		X	X		X								
WMDs	X	X	X			X	X				X			X			
Terrorism		X	X			X	X		X		X	X		X	X		
Attack on computer networks									X								
Fragile states			X			X								X	X	X	
Crime						X			X					X			
Action																	
Capability improvement of Armed Forces		X	X		X	X	X	X	X	X	X	X		X		X	
Technological innovation								X			X	X	X				
Cooperation	X	X			X		X	X	X	X	X			X		X	X
Humanitarian/Peace			X		X	X	X	X	X	X	X	X	X	X	X		X
"Daily" tasks		X			X	X			X	X	X			X	X		
Diplomacy		X				X								X			
Image improvement					X	X											X
Non-proliferation		X				X	X		X					X			
Who																	
Unilateral																	
Citizens/People				X	X	X	X	X	X		X			X		X	
Government	X	X	X		X	X	X		X	X	X	X		X		X	X
Defense apparatus		X			X		X		X	X	X	X		X		X	
Nation					X	X	X	X	X	X	X	X	X	X	X	X	
Bilateral																	
Africa					X	X		X	X								
Latin America																	
United States			X		X	X			X					X	X		
Other countries						X		X		X							
Multilateral																	
Neighbors	X																
Allies				X	X		X							X			X

Ambition	AUS (2000)	AUS (2003)	AUS (2005)	AUS (2007)	B (2003)	B (2008)	DK (2004)	F (2003)	F (2008)	NL (2000)	NL (2003)	NL (2005)	NL (2007)	UK (1998)	UK (2003)	UK (2008)	WFP (2004)
EU					X	X	X	X	X	X	X	X	X	X	X	X	
UN	X				X	X	X			X		X	X	X	X	X	X
NATO					X	X	X	X	X	X	X	X	X	X	X	X	
OSCE					X					X				X			
ESDP						X	X			X					X		
International Community	X	X			X	X				X	X	X	X	X	X		X
Humanitarian																	
Civil-Military						X	X					X	X			X	X
Where																	
National																	
Home Security					X	X	X	X	X	X	X	X	X	X	X	X	
National Sovereignty				X			X	X	X	X	X	X	X				
Overseas Territories								X	X	X	X	X	X	X	X	X	
Citizens abroad						X		X	X			X		X	X		
International																	
International		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Space									X								
When																	
Focus																	
Short Term								X		X	X	X	X			X	X
Long Term	X		X		X	X	X	X	X	X	X	X		X		X	X
Action																	
Anticipation									X							X	X
Prevention		X	X		X	X	X	X	X					X		X	X
Respond		X	X				X		X					X			X
Conflict management					X	X					X	X					
Intervention						X			X						X		
Reconstruction						X			X		X						

Table 4 exemplifies non-normative benchmarking that might still be useful to various countries. It is a systematic data-driven comparison. There is no "right" or "wrong" in this table, no "better" or "worse." And yet any country working on a new white paper might benefit from such an overview to double-check whether it has covered all its bases. For instance, if a number of new items in this table appear in the high-level policy documents of most other friendly countries, but not in one's own – drafters of such policy documents might bring this to the attention of their political leadership. They may ultimately still decide to include or exclude certain elements, but at least such a synoptic overview might trigger useful discussions about such issues.

Level of Ambition: The HCSS Audax Index

The second aspect of the ambition level we want to illustrate here is the "gutsiness" of a country's defense ambition as expressed in its highest level policy documents. Again we used the topic-to-metric decomposition method and disassembled the very abstract concept of level of ambition into a number of indicators that we could actually operationalize. The HCSS Audax Index thus aims to represents an overall view of a referent's total (stated) defense ambition and is based on the following six indicators:

1. **Reach:** the explicit mentioning of the geographical expanse within which the country is willing to take military action.
2. **Concurrency:** the amount of operations a country is willing to engage in simultaneously (normalized for the size of the country).
3. **Interoperability:** the degree to which countries are willing to remain interoperable with other (militarily more capable) nations (like the U.S. or the UK).
4. **Unilateralism:** the level of international agreement needed to justify military action (i.e., is a United Nations mandate explicitly required for military action or not).
5. **Pre-emption:** the willingness to resort to pre-emptive military action in order to counter possible developing threats.
6. **Violence spectrum:** the explicit mention of the level of violence with which the country is willing to operate (e.g., explicitly also in the highest regions of the violence spectrum or not).

These radar charts represent the values of these parameters for each country as coded (by HCSS) on the basis of the aforementioned documents. To give a notional but concrete example: a country with a totally "full" radar chart would be a country that is willing to send troops all over the globe in a number of concurrent operations engaging, if necessary, even preemptively and at the highest levels of violence and without a UN mandate and while remaining fully interoperable at the highest levels with the United States.

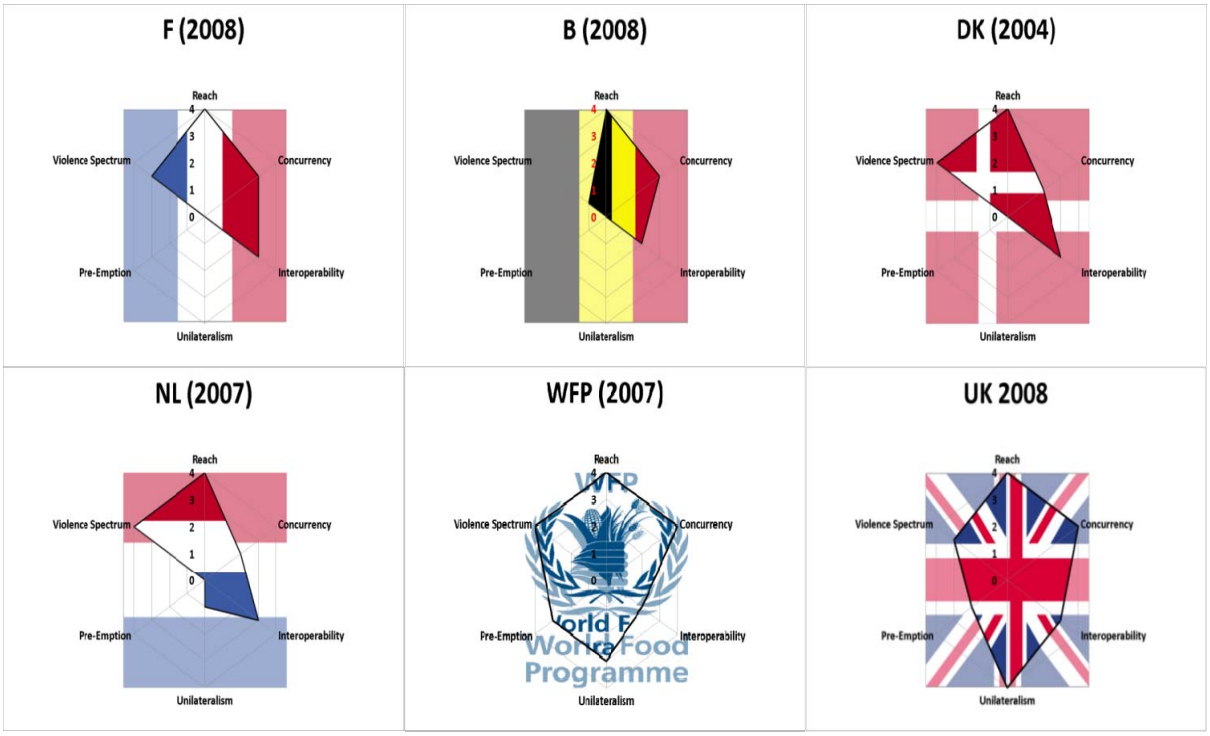

Figure 21. The HCSS Audax Index

One immediate observation that emerges from a comparison of the various radar charts is that both Australia and the UK score significantly higher on unilateralism and pre-emption. Visually, this is illustrated by the skewed graphs of France, Belgium, and Denmark and the rounder graphs of Australia and the United Kingdom. This distinction between the two Anglo-Saxon countries and the others is interesting because there we shall see a similar divide in the logic of their capability development processes.

When we look at the radar charts, we note that all of the countries score high on the Reach parameter. This represents a big change for the European referents which were reluctant to engage "out of area" at the end of the Cold War. The charts show that this reluctance has now been overcome, at least in these countries' strategic thinking. Only Australia scores a 3 whereas the rest scores the maximum of 4. This illustrates the commonly shared (post-September 11) assumption that threats have become globalized and that events in one region have spill-over effects elsewhere. A common theme therefore in all the high-level documents under review is that the countries' interests benefit from a more stable and secure world. It will be interesting to observe to which extent this global focus will withstand the possible consequences of the current financial economic crisis.

Scenarios

Scenarios are used to help referents operationalize the strategic environment within which they may have to operate in the future. Consequently, scenarios provide the context for capabilities-based planning and are an integral part for the remainder of the capability analysis process, being referenced and reused throughout the process. We examined the use of scenarios with respect to the number of scenarios used, their degree of specificity, and how pivotal their role is in each referent. Because scenarios (or in broader terms, foresight) plays an essential role in capability generation, their robustness and capacity to adequately inform defense planners warrants closer examination.

Number of scenarios used

This slidebar in Figure 22 measures the number of scenarios used in each defense planning cycle. The number of scenarios may be related to their degree of specificity, and—by extension—to how robust they are in handling uncertainty in the strategic environment.

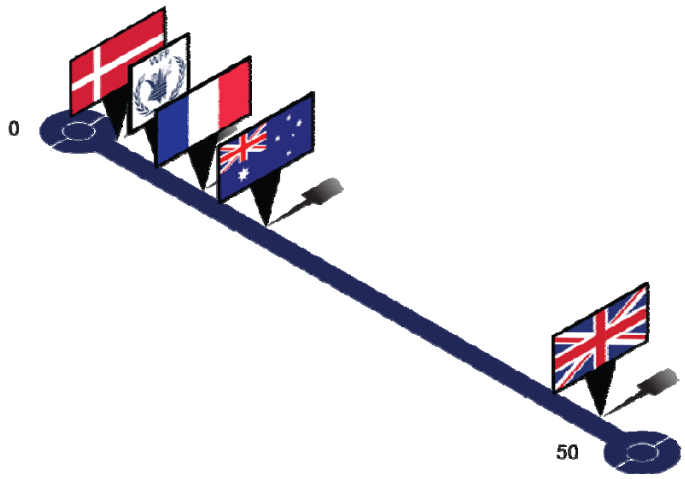

Figure 22. Number of Scenarios Used

Of the referents under review here, the UK makes the most use of scenarios by far. In the biannual Defense Strategic Guidance exercise, UK defense planners develop and run 46 scenarios. The Australian Defense Force typically develops approximately 10 Illustrative Planning Scenarios per year. These are used at the highest level of defense planning to map the long-term strategic environment. While there are only 10 Australian Illustrative Planning Scenarios (AIPS), a multitude of operational scenarios are also used for specific operational planning. Information on France is sketchy on this point, but there seems to be less emphasis on scenarios and more on broader geostrategic analysis. From the limited

material available, it appears that Denmark makes no use of scenarios in informing their capability analysis process. There is no predetermined number of scenarios the WFP uses. Rather, scenarios are constructed on an ad hoc basis as part and parcel of the vulnerability assessment phase in Emergency Food Security Assessment.

Specificity of scenarios

The slidebar in Figure 23 represents an interpretation of the degree of specificity in the scenarios used to facilitate the capability analysis process. Ideally, scenarios should cover the full spectrum of plausible threats. A wider set of scenarios is increasingly seen as a better guarantee for capabilities that are more robust against future shocks. At the same time, a highly specific set of scenarios (point scenarios) is also increasingly seen as vulnerable to unforeseen shifts in the strategic landscape. The problem here is that often the highly specific scenarios that are used for operational (or short-term contingency) planning are "dual-used" as long-term scenarios for forward defense planning. This allows military planners, who tend to be much more familiar (and comfortable) with operational planning than with forward planning, to fall back on existing planning "investments" that typically suffer from excessive "presentism." Succumbing to the temptation of turning forward defense planning into a form of glorified operational planning, however, means that typically insufficient uncertainty is built into the scenarios, thus leading to suboptimal capability choices over time.

To deal with the "point scenario" problem, some key countries are building in "shocks" or "branches" around their existing scenario set; we clearly are seeing a trend towards more parameterized approaches to foresight.

AIPS represent the highest level of scenarios use in defense planning. Due to their broad strategic outlook and long time horizon (15 to 25 years) AIPS tend to be parameterized. More specific operational scenarios are developed at the command level to plan specific operational campaigns. The UK scenarios are at the campaign level, taking in account the contributions of allies and played out in different time epochs.

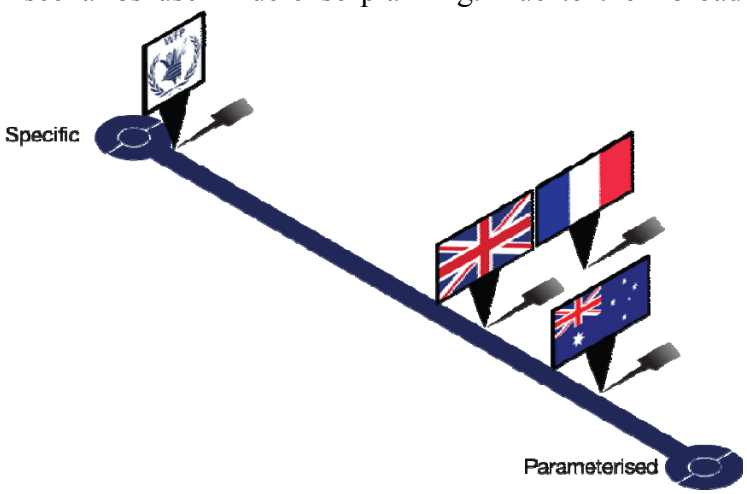

Figure 23. Specificity of Scenarios

WFP scenarios are limited to exploring the effects of market shocks on food consumption rates for various groups on people, and are used as a vulnerability assessment tool, not necessarily as a dedicated input to capability generation.

Overall process

Capability analysis is a complex undertaking that can be looked at from a variety of different perspectives. It can be analyzed (and benchmarked) from an *institutional perspective:* as an allocation of responsibilities to bureaucratic agencies. It can also be viewed as a series of sequential steps taken to get from point A to B (*process-based perspective*). Our description (and benchmarking) of the capability analysis efforts of the referents is primarily focused on the underlying functional logic (*functional perspective*) of the process, which functional tasks

the referents execute in order to translate the higher-level policy guidance into a set of defense capabilities. This chapter will thus attempt to describe the main underlying logic of modern-day capability analysis with its various functional building blocks.

Understanding the Z-Charts

Given the differences in organizational structures and processes between referents, we present the capability analysis process by dissecting it into a number of key generic functional "building blocks" that can be found back in all (or at least most) referents. We present these main building blocks as anchor points in a Z-shaped diagram we call a "Z-Chart."

The Z-Chart represents our notional reconceptualization of the capability generation process in each referent. Read from the top left to the bottom right, it follows the process along three main lines, with the turning points signaling a transition from one stage to another. Although depicted as a linear path from the reception of High Level Policy Parameters to a Capability Plan, the actual processes themselves need not be, nor should they be viewed as, purely linear. All capability generation schemes are channeled through an intricate bureaucratic machinery that goes through a multitude of processes and sub-processes (often simultaneously and/or iterated) and is sometimes redirected as the strategic environment dictates.

Stage 1

The first line starting at the top left represents the effort to translate the high-level policy guidance coming down from the highest levels of political leadership into a corresponding set of capability requirements. Generally speaking, this stage remains quite opaque for reasons of both methodological complexity and national security-related sensitivities. There is, however, a clear commonality in the actions taken and the concepts that emerge as the referent's process unfolds from one pole end to another. At the same time, the exact shape, sequencing, and impact of these various elements will vary from organization to organization.

The first step in this first stage is the translation of the high-level policy parameters into a set of more concrete planning assumptions that defense planners can actually work with. These planning assumptions specify areas like the types of missions and the scale and level of concurrency. Given the quite abstract and sometimes nebulous nature of many higher-level policy documents (especially for national security), this translation process is far from trivial, and requires close interaction between the more "political-military" parts of the defense organizations and their more "military-technical" and operational counterparts. High-level documents, for instance, will often stipulate that defense organizations have to be able to cover a number of threats without specifying exactly how many of such contingencies their armed forces are supposed to be able to cover simultaneously. Defense planners argue that without such specifications, it is practically impossible to answer the essential question, "How much is enough?" Defense planning assumptions (which vary in shape, scope, and across the referents) are therefore typically found in separate (and usually classified) documents.

On the basis of these defense planning assumptions, defense planners use a number of different analytical building blocks to "engineer" capability packages. These include (and many of them re-occur in subsequent stages of capability generation):

- **Scenarios** are used to help referents operationalize the strategic environment within which may have to operate in the future. This environment will usually be described in the higher-level documents, but typically at a level of abstraction that makes deriving concrete capability choices from these threats difficult, if not impossible. Mandating that a referent has to be able to execute a certain number of peace support operations in failed or failing states, for instance, says little about parameters such as terrain, climate, distance, permissiveness of the security environment, alliance partners, or degree of host nation support. Yet these are precisely the critical planning parameters that are required for making concrete choices (for operational planning and, in the mind of most defense planners, also for forward defense planning) because only they can guide decisions on the types of strategic or tactical mobility, on force protection, etc. Therefore, defense planners typically develop a set of more detailed planning scenarios that will embody some additional concrete situation-specific planning assumptions they feel are required to make informed and robust choices. Scenarios thus become a vital input in identifying capability strengths and weaknesses, and may aid a whole-of-force capability balance-of-investment[76]. The inputs, degree of specificity, and the exact narrative of the scenarios are increasingly bolstered by modeling, simulation, and scientific experimentation by and/or with the defense analytical community.
- **Partition schemes**. Military capabilities—and *a fortiori* defense or security capabilities—span an extremely broad (and, as nations start moving towards more comprehensive security planning approaches, increasingly broadening) array. To manage this complexity, various referents use different partition schemes to cut up the larger area of defense (or security) capabilities into more manageable subareas. Traditionally, this was done essentially along the lines of the different operational environments (air, land, sea) as embodied in the services. While still of great importance, it is increasingly recognized in all examined countries that the environment-based partition scheme, and the stovepiping that results from it, leads to a number of dysfunctional consequences (like duplication, "holes," lack of interoperability, etc.) We have therefore seen a number of more *functional* partition schemes emerge to either complement or even replace the *service-based* one.
- **Time horizons**. The time horizon of defense organizations is unusually long in comparison with most other government departments and even—with the possible exception of highly capital-intensive industries such as the petrochemical sector—with the private sector. This means that, just as with the partition schemes for "capability" as such, defense organizations also have to break down the 20+ year time horizon into more manageable "epochs" (e.g., priorities for the first 5 years, for the subsequent 10 years, and for beyond that). As with any partition scheme, this creates seams (e.g., tensions between short-term capability priorities and medium-term ones) that different countries address in different ways (and with differing degrees of success).

[76] By this we mean a trade-off analysis of the benefits and consequences of prioritizing one capability platform at the expense of another in a resource-constrained environment.

- **Operational concepts**. In the last decade, the larger (at least Anglo-Saxon) countries have also added "concepts of operations" (also called "operational concepts") to the analytical suite they use to translate policy into capability requirements. The thinking behind this addition is that before any scenario can be translated into capability requirements, one would like to have an idea about **how** the challenges in that scenario can be addressed. These concepts come in various forms and shapes and are used at different levels in different referents. An (early) example, for instance, is the concept of network-enabled capabilities. Defense concepts like these seldom develop in a vacuum and often arise from the interplay between scenarios, scientific experimentation and validation, and military judgment.
- **Military judgment.** Despite the emergence of various analytical support tools for defense planning, the role of military judgment remains central. All participants in the process remain acutely aware of the various limitations of the existing suite of software-based support tools. This means that in the final analysis, the experiences and intuitions of the uniformed military (but increasingly also of non-military operators and experts) remain central to ensure the integrity and the quality of the entire process.
- **Operational analysis.** Scientific support to defense planning has increased significantly in size and scope in the past decades—including in the translation from policy to capability requirements. This manifests itself in various analytical support software tools that increasingly try to crystallize expert judgment, scientific knowledge, and empirically validated findings into traceable tools that can help elucidate some of the key choices to be made in the process.
- **Industry input.** Depending on the referent, contact with the defense industrial community will start either sooner or later in this stage, especially when scenarios identify a deficiency entailing a significant technological or acquisition dimension. Furthermore, the defense technical research community may also rely on data from the defense industry in the course of validating scenario mathematical models, narratives, and outputs and to aid a whole of force capability balance-of-investment.

These building blocks are assembled by the various referents into a set of ***capability requirements***—capabilities that are derived from the higher-level policy guidance by means of the analysis carried out (with the help of the building blocks) in Stage 1.

Stage 2

Stage 2 entails a referent's attempt to funnel a (typically broad) array of capability requirements into a coherent set of capability packages that have been audited against baseline capabilities (capabilities that either already exist or are in the pipeline). In most referents, this stage will include the translation of the capability requirements into concrete capability goals for each element of the prevailing partition scheme. Typically, this generates a set of capability shortfalls that will then have to be remedied on the basis of some additional analysis that will take place in Stage 3.

This stage ends when the referents conduct an internal assessment (i.e., an "audit" of the capability packages stemming from the judgments rendered on the first axis). A capability audit represents a form of "health check" without recommendations, that is, it tells you what will happen if nothing is changed or how well the currently planned force will meet the goals.

Subsequent balance-of-investment studies will then inform you about what you can actually afford to fix in Stage 3. The audit was introduced to replace a system where managers only looked for gaps to justify increased investment. The audit forced them to acknowledge where they were strong and where they had surplus.[77] Should the referent have a stand-alone capability generation group, its most intense efforts will probably gravitate towards conducting such an audit.

In reality, the development of concepts and of specific capability options may occur with significant overlap. This is why in many of the referents we observe a reoccurrence throughout the various stages of scientific experimentation or scenario work, with much attention being given to ensure that the capability packages proposed are in line with certain defense concepts the referent wants to adhere to from start to finish. Typically these concept development plans are known to as "roadmaps." The audit may also include an examination of interoperability issues depending on the primacy the organization places on various strategic partnerships.

Stage 3

The final axis on the capability generation path is marked by the capstone output—a specific capability generation plan (for countries typically the defense plan) that outlines what, when, and how much of each capability option will be implemented (and procured). At this point, the options will be clearly articulated and the scope of the endeavor will be narrowed down considerably.

In this last stage of the capability generation process a number of different (but highly interconnected) tools are increasingly being used:

- **Capability investigations**—Once a capability shortfall has been identified on the basis of Stages 1 and 2, there may still remain various different options to fill that capability shortfall from a purely operational point of view. For example, if strategic lift is identified as a critical shortfall (as it has within the NATO Alliance for well over a decade), defense planners will still have to investigate the various options available for this (e.g., whether to buy it, lease it, or invest in "real options"; whether to go for airlift or sealift; and which options to go for within air lift). The trade-off analysis between these various capability options lies at the heart of these "capability investigations," which focus primarily on optimal operational effectiveness.
- **Balance-of-Investment studies—**Many defense organizations are also increasingly starting to factor in value-for-money considerations in their capability generation processes. Money has always been an important consideration in defense planning, but recent cost trends, spectacular cost overruns, shrinking defense budgets, and a general increased emphasis on government performance management have made the financial dimension more imperative than ever. We increasingly see Balance-of-Investment studies appearing at the level of individual capabilities (especially for the high-ticket items), but still see little publicly available evidence of it at the macro-level (e.g., whether one gets more overall "defense value for money" from fighter jets

[77] We are indebted to Dr. Ben Taylor from DRDC-Canada for this insight.

or command, control, communications, computers, intelligence, surveillance and reconnaissance assets).

- **Risk management**—Recent experiences with cost overruns or the acquisition of suboptimal capabilities have honed our defense organizations' interest in and sensitivity to risk analysis. Even if a referent has succeeded in identifying the optimal option for addressing a capability shortfall from an operational effectiveness point of view **and** from a value-for-money point of view, there may be a number of risk factors that may make another option preferable. As with balance-of-investment studies, we are increasingly finding these considerations at the program-level, or even within some of the partition elements (e.g., capability sub-areas such as "mobility"), but much less so at the macro-level (e.g., risk management for major technological disruptions).

After these analyses, all that remains is to reassemble the various capability packages into an overall defense capability plan. This requires close coordination with the defense industrial community, and it is here that the building block icon of industry makes a universal appearance, as exhibited in Figure 24. The process concludes with an annual performance assessment designed to measure the effectiveness of the referent in achieving its capability objectives within the mandates and confines of the High Level Policy Guidance. In essence, closing a strategic "sense and response" feedback loop, this assessment has its own systems and methodologies, known as performance management

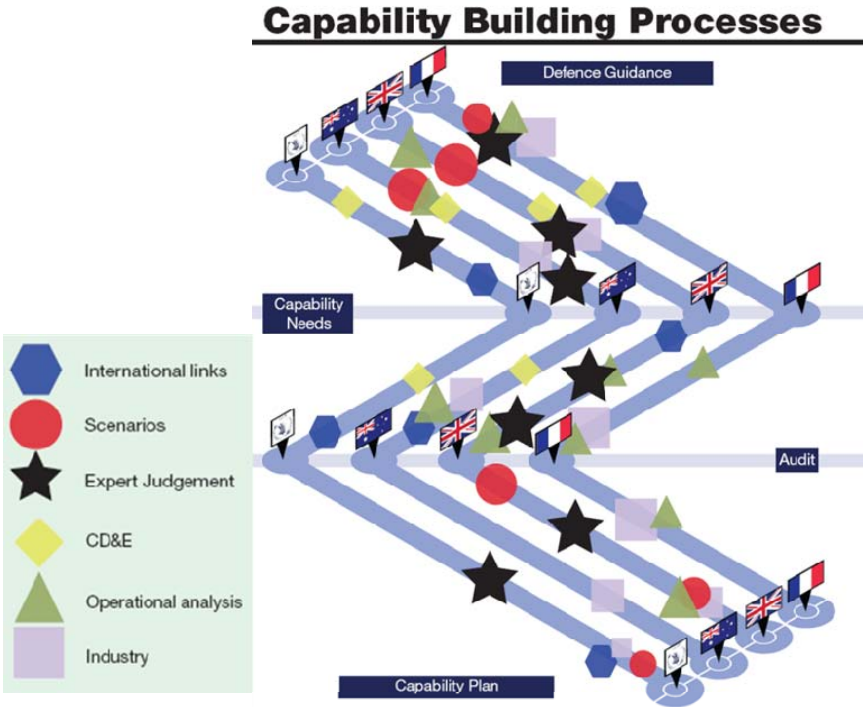

Figure 24. The HCSS Z-Chart: Capability Building Process

Impact of the study

This benchmarking study (of which we only presented some examples in this paper) led to a number of intense discussions between the HCSS team that executed the study and a number of high-level MoD participants in the large bottom-up defense review that was being conducted in parallel. HCSS identified a number of concrete nuggets from the benchmark study that were discussed in these meetings (see Figure 25).

A few of these nuggets are now being implemented within the Netherlands Defense Organization, including the basic idea that the organization should be able to close the "strategic defense management" loop. The creation of a new entity with responsibility for the department's strategy, knowledge, and innovation agenda within the organization can also be attributed to this evidence-based systematic comparison.

	A	B	C	D	E	F	G	H	I
10	Points of Interest	Buy-in	Transparency of outcome	Transparency of process	Scale of effort	Traceability	Adaptiveness	Description/Example	Significance
11	Process								
13	Some countries are committed to stating resource parameters within a White Paper, and then reporting the status of Defence's success in allocating funds against the parameters established in special sections of subsequent White Papers.							Australia has a special section devoted to reporting resource allocation according to previous White Paper statements.	Doing so continually reasserts an emphasis on A) identify a single 'well-spring' of strategic guidance and B) demonstrating a commitment to closing the loop at the highest levels of Defence.
14	A formal process for societal 'buy-in' prior to the Defence White Paper publication.							Australia initiated Community Consultation Panels as part of a series of Companion Reviews to augment the publication of a new White Paper.	Represents a more comprehensive approach to stakeholder analysis and serves to strengthen the 'social contract' between the tax payer and the service provider.
15	Some referents use long term Coalition Agreements to stabilize funding parameters and ambition statements over time.							Denmark combines specific resource allocation projects into its White Paper (Defence Agreement) over a comparatively long time line (5 yrs). However, the Minister of Defence submits an annual status report to the Parliamentary Defence Committee as to the implementation of the objectives stated in the Defence Agreement.	The presence of coasted, long range objectives embedded in a referent's White Paper may signify a 'micro SDM loop' that occurs above the Defence Guidance portion of the SDM loop as a whole.
18	The deliberate synchronization of formal 'Strategy Reviews' and formal budget reviews.							The United Kingdom deliberately synchronizes formal reviews of its 20 year Defence Strategy in tandem with the 4 year Defence Spending Reviews.	Perhaps representative of the growing awareness for the need not to isolate funding guidance from strategic guidance as the two are part and parcel to the accomplishment of the highest level strategic objectives.
20	Some referents have adopted a multi-faceted budgetary formulation and reporting structure as opposed to a simple planned vs. actual defence expenditure reports.							For example, the United Kingdom conducts comprehensive 4 year Spending Reviews (in tandem with a review on the implementation of Defence Strategy) then bases the Annual Expenditure Plan upon the findings of the Spending Reviews. Running parallel to the forward-looking Expenditure Plan, is the retrospective Annual Report which assesses the previous year's Expenditure Plan to deliver the services specified in the Public Service Agreements.	Just as Defence organisations have embraced macro-level Strategic Reviews on whether the organisation is appropriately orientated for the long term, so to has the financial sector of Defence.

POI Overview / POI Distribution / Legend

Figure 25. "Nuggets" Distilled from the "Closing the Loop" Benchmark Study

Defense Benchmarking: Where Do We Stand?

The two previous sections described how benchmarking is currently done in both the private sector and the public sector. They pointed out how popular benchmarking has become in the private sector and how there is a solid consensus about the benefits (both perceived and demonstrable) benchmarking has brought there, not as a panacea, but as one of many useful tools that can be used to improve an organization's performance. We also explained how in the public sector, international organizations like the OECD are increasingly playing the role of trusted collectors and curators of different insights culled from various benchmarking efforts that are then used by national governments to adjust their own policies in light of those findings.

This section has presented two very different, but interestingly complementary extant approaches to defense benchmarking. It has described the experiences of one individual country, the Netherlands, which has started using (mostly descriptive) defense benchmarking more systematically for its own planning purposes on various ad-hoc issues. This section of the paper has also presented a large one-of-a-kind study completed by a private consultancy with a (mostly normative) large international benchmark comparing the relative performance of 33 of the most advanced defense organizations on a number of important parameters.

We want to emphasize how these two examples—which we see as best of kind—remain far from ideal. Taking our cue from the work of an organization like the OECD in areas like public health or education, we cannot but be surprised that there is at present not a single public international effort to systematically compare the experiences various defense organizations are accumulating on providing defense value for money. But both the McKinsey defense benchmarking study and the 15+ defense benchmarking studies that were completed in the Netherlands Defense Organization illustrate that there is enough publicly available information to come—with a healthy dose of creative rigor—to meaningful comparisons that can be used by defense organizations to improve their performance. They also show how much work still has to be done to collect all of those data and to make them reliably (and traceably) comparable.

We are confident that national efforts (both unilateral and minilateral[78]) to learn from others in the defense and security area will continue. We also surmise that consultancies will continue to build up and exploit their own proprietary knowledge bases with the comparative insights they glean from the work they do for various defense organizations across the world. Defense organizations are likely to benefit from both of these efforts and it might even be useful to explore ways to come to some form of public-private partnership between them. But currently we still feel a preferable model would be for some international organization to assume this task of a clearinghouse of evidence-based benchmarking efforts to the benefit of its member states—along the lines of the work that the OECD does in other policy areas.

CONCLUSION

The battle for better capabilities is a critically important one—for the Alliance, for its individual member states, and arguably even for international security. Demand for the public goods of international security and stability remains high. Their supply remains distinctly

[78] Scandinavian countries, for instance, have accumulated interesting experiences with benchmarking various aspects of their defense planning processes with each other.

suboptimal. The North Atlantic alliance of liberal democracies continues to aspire to a unique role in bridging the gap between demand and supply for international security and stability. But the capabilities that are required for successfully fulfilling this role become ever more difficult to generate and sustain. For better or worse, capabilities remain overwhelmingly national: they are born and grown nationally through national defense planning processes over which outsiders (including international organizations like NATO) have little sway.

NATO's efforts to influence national capability efforts have focused primarily on the employment (downstream) stage of the life-cycle of capabilities and have left the upstream almost entirely to the nation states.

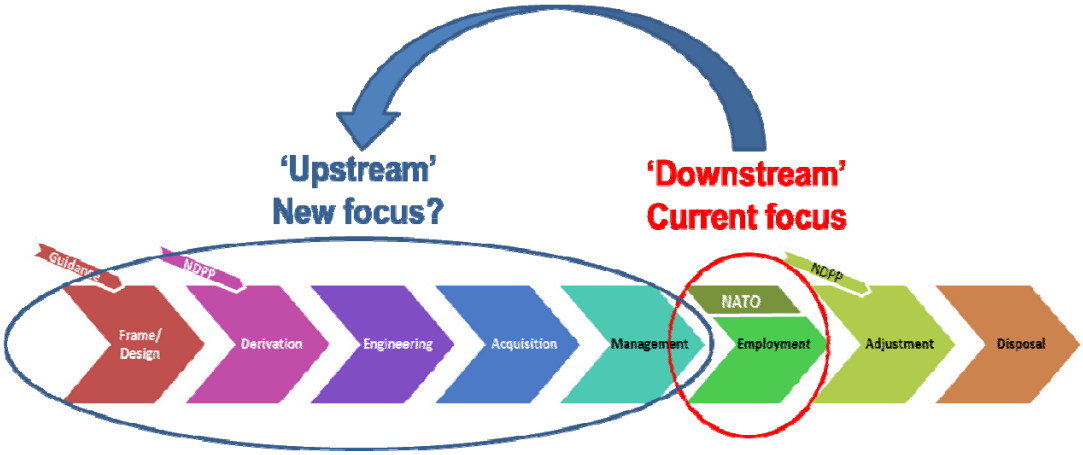

Figure 26. Taking the Focus Upstream

The current (geo)political, technological and especially financial realities require NATO to take the battle for capabilities upstream. National defense planning processes are one of the most complex planning endeavors on this planet and all NATO nations—even the bigger ones—are struggling with it. There is ample room for improvement through learning from others throughout the capability life cycle and NATO is ideally (and uniquely) positioned to advance this learning mechanism. At each step in this chevron-chart every individual country makes myriad decisions—big and small—that determine the ultimate force that becomes the pool from which nations apportion forces to NATO. Many of these choices are currently not systematically mapped by any national or international instance. Yet this paper argues that every individual country and the alliance as a whole would really benefit from more comparative insights into what does or does not work in the upstream capability development and management stages.

More and more defense organizations today produce ever larger quantities of publicly available (and approved) data and documents—primarily for their own domestic audiences (accounting chambers, parliaments, publics, but also for educational purposes). These datasets and documents represent a burgeoning treasure trove that can be mined for evidence-based comparative analysis, which in turn can inform and inspire national defense planning processes. This paper has provided some concrete examples of the results and the types of insights that such benchmarking efforts can yield. It has also emphasized that there remain many hurdles to be overcome. Efforts by individual (or small groups of) nations, companies or think tanks can certainly provide valuable inputs that can be used by decisionmakers across the Alliance (provided they are made publicly available, preferably in English). But they are unlikely to singlehandedly be able to overcome the various hurdles (also analytical) that rigorous defense benchmarking requires. To be truly effective, defense benchmarking is in need of a higher-level catalyst as a strategic engine. NATO—and particularly its Allied

Command Transformation, the Alliance's leading agent for change—is ideally placed for such a role. It has the mandate, the authority, and the resources to build up a more systematic benchmarking facility within the Alliance. Such an effort is consistent with "driving, facilitating, and advocating continuous improvement of Alliance capabilities to maintain and enhance the military relevance and effectiveness of the Alliance." The knowledge base such a facility would produce could be put at the benefit of national defense planners, thus taking the battle for better capabilities upstream. In this way, defense benchmarking could become a new tool in a richer and smarter strategic defense management toolbox in line with what NATO's new push for smart defense is trying to achieve.

BIBLIOGRAPHY

"About OECD", n.d. http://www.oecd.org/pages/0,3417,en_36734052_36734103_1_1_1_1_1,00.html.

Adebanjo, D., C. Street, R. Mann, and A. Abbas. "Twenty-five Years Later–a Global Survey of the Adoption and Implementation of Benchmarking: Management Fad or Established Operations Management Tool" (2009).

Adebanjo, Dotun, Ahmed Abbas, and Robin Mann. "An Investigation of the Adoption and Implementation of Benchmarking." *International Journal of Operations & Production Management* 30, no. 11 (2010): 1140–1169.

Adebanjo, Dotun, Robin Mann, and Ahmed Abbas. "Benchmarking - BPIR.com." *BPIR Management Brief* 4, no. 5 (n.d.). http://www.bpir.com/benchmarking-bpir.com/menu-id-73.html.

Ahmed, P. K, and M. Rafiq. "Integrated Benchmarking: a Holistic Examination of Select Techniques for Benchmarking Analysis." *Benchmarking: An International Journal* 5, no. 3 (1998): 225–242.

Ammons, David N. "A Proper Mentality for Benchmarking." *Public Administration Review* 59, no. 2 (March 1999): 105–109.

Anand, G., and R. Kodali. "Benchmarking the Benchmarking Models." *Benchmarking: An International Journal* 15, no. 3 (2008): 257–291.

Andersen, Bjørn, and Per-Gaute Pettersen. *The Benchmarking Handbook: Step-by-step Instructions*. Springer, 1996.

Anderson, K., and R. McAdam. "An Empirical Analysis of Lead Benchmarking and Performance Measurement: Guidance for Qualitative Research." *International Journal of Quality & Reliability Management* 22, no. 4 (2005): 354–375.

———. "Reconceptualising Benchmarking Development in UK Organisations: The Effects of Size and Sector." *International Journal of Productivity and Performance Management* 56, no. 7 (2007): 538–558.

Auluck, Randhir. "Benchmarking: A Tool for Facilitating Organizational Learning?" *Public Administration & Development* 22, no. 2 (May 2002): 109–122.

Beinhocker, Eric. The Origin of Wealth□: Evolution, Complexity, and the Radical Remaking of Economics. Boston Mass.: Harvard Business School Press, 2006.

Berg-Knutsen, Espen, and Pål Remy Østbye. "Economic Analysis at FFI", October 2005. http://www.ffi.no/no/Publikasjoner/Documents/Economic_analysis_at_FFI.pdf.

"Big Savings from Little Things: Non-Equipment Procurement." *McKinsey on Government. Special Issue: Defence*, March 2010.

Bohannon, John. "Google Books, Wikipedia, and the Future of Culturomics." *Science* 331, no. 6014 (January 14, 2011): 135.

Bowerman, Mary, Graham Francis, Amanda Ball, and Jackie Fry. "The Evolution of Benchmarking in UK Local Authorities." *Benchmarking: An International Journal* 9, no. 5 (2002): 429–449.

Braadbaart, Okke, and Benni Yusnandarshah. "Public Sector Benchmarking: a Survey of Scientific Articles, 1990—2005." *International Review of Administrative Sciences* 74, no. 3 (2008): 421 –433.

Bullivant, J. R.N. *Benchmarking for Continuous Improvement in the Public Sector*. Longman Information & Reference, Harlow, 1994.

Camp, Robert C. Benchmarking: The Search for Industry Best Practices That Lead to Superior Performance. Taylor and Francis, 2006.

Cowper, J., and M. Samuels. "Performance Benchmarking in the Public Sector: The United Kingdom Experience." *Benchmarking, Evaluation and Strategic Management in the Public Sector* (1997): 11–32.

Dattakumar, R., and R. Jagadeesh. "A Review of Literature on Benchmarking." *Benchmarking: An International Journal* 10, no. 3 (2003): 176–209.

Dolowitz, D., and D. Marsh. "Who Learns What from Whom: a Review of the Policy Transfer Literature." *Political Studies* 44, no. 2 (1996): 343–357.

Dolowitz, David P, and David Marsh. "Learning from Abroad: The Role of Policy Transfer in Contemporary Policy-Making." *Governance* 13, no. 1 (January 1, 2000): 5–23.

Dorsch, Jeffrey J., and Mahmoud M. Yasin. "A Framework for Benchmarking in the Public Sector: Literature Review and Directions for Future Research." *International Journal of Public Sector Management* 11, no. 2/3 (1998): 91–115.

"European Benchmarking Network", n.d. http://circa.europa.eu/irc/euradmin/eubenchmarking/info/data/en/ebnsite/page1b.htm.

Fernandez, P., I. P McCarthy, and T. Rakotobe-Joel. "An Evolutionary Approach to Benchmarking." *Benchmarking: An International Journal* 8, no. 4 (2001): 281–305.

Flynn, N. *Public Sector Management*. Sage Publications Ltd, 2007.

Fong, S. W, E. W.L Cheng, and D. C.K Ho. "Benchmarking: a General Reading for Management Practitioners." *Management Decision* 36, no. 6 (1998): 407–418.

Francis, G., and J. Holloway. "What Have We Learned? Themes from the Literature on Best-practice Benchmarking." *International Journal of Management Reviews* 9, no. 3 (2007): 171.

Ginn, D., and M. Zairi. "Best Practice QFD Application: An Internal/external Benchmarking Approach Based on Ford Motors' Experience." *International Journal of Quality & Reliability Management* 22, no. 1 (2005): 38–58.

Gomes-Casseres, B. "Competing in Constellations: The Case of Fuji Xerox." *Strategy+ Business, First Quarter* (1997).

Van Helden, G. J, and S. Tillema. "In Search of a Benchmarking Theory for the Public Sector." *Financial Accountability & Management* 21, no. 3 (2005): 337–361.

Hewitt, Rachel. Map of a Nation: A Biography of the Ordnance Survey. Granta Books, 2011.

Hinton, M., G. Francis, and J. Holloway. "Best Practice Benchmarking in the UK." *Benchmarking: An International Journal* 7, no. 1 (2000): 52–61.

Hood, Christopher, Ruth Dixon, and Craig Beeston. "Rating the Rankings: Assessing International Rankings of Public Service Performance." *International Public Management Journal* 11, no. 3 (December 2008): 298–328.

———. "Rating the Rankings: Assessing International Rankings of Public Service Performance." *International Public Management Journal* 11, no. 3 (December 2008): 298–328.

"IRRIIS – Integrated Risk Reduction of Information-based Infrastructure Systems", n.d. http://www.irriis.org/.

Jarrar, Y. F, and M. Zairi. “Future Trends in Benchmarking for Competitive Advantage: a Global Survey.” *Total Quality Management & Business Excellence* 12, no. 7 (2001): 906–912.

Jarrar, Yasar, and Giovanni Schiuma. “Measuring Performance in the Public Sector: Challenges and Trends.” *Measuring Business Excellence* 11, no. 4 (2007): 4–8.

Katzman, Kenneth, and Congressional Research Service. *Iraq: Politics, Elections and Benchmarks*. DIANE Publishing, n.d. http://fulltextreports.com/2011/05/10/crs-iraq-politics-elections-and-benchmarks/.

Knill, C. “Introduction: Cross-national Policy Convergence: Concepts, Approaches and Explanatory Factors.” *Journal of European Public Policy* 12, no. 5 (2005): 764–774.

Kugler, Richard L. *Laying the Foundations: The Evolution of NATO in the 1950s*. Santa Monica, CA: RAND Corporation, 1990.

Kyrö, P. “Revising the Concept and Forms of Benchmarking.” *Benchmarking: An International Journal* 10, no. 3 (2003): 210–225.

“Lakes Guides, Bench Marks, Cumbria, Frameset”, n.d. http://www.geog.port.ac.uk/webmap/thelakes/html/topics/benmrkf.htm.

“Lessons from Around the World: Benchmarking Performance in Defense.” *McKinsey on Government. Special Issue: Defence*, March 2010.

Lundvall, Bengt-Åke, and Mark Tomlinson. “International Benchmarking as a Policy Learning Tool.” In *Benchmarking in the Public and Nonprofit Sectors: Best Practices for Achieving Performance Breakthroughs*, by P. Keehley and N. N Abercrombie, 203–230. Jossey-Bass Inc Pub, 2008.

Magd, H., and A. Curry. “Benchmarking: Achieving Best Value in Public-sector Organisations.” *Benchmarking: An International Journal* 10, no. 3 (2003): 261–286.

Malik, Khaleel, and Paul Cunningham. “Transnational Policy Learning in Europe: Attempts to Transfer Innovation Policy Practices.” *Innovation: Management, Policy & Practice* 8, no. 3 (July 2006): 262–272.

“Mastering Military Maintenance.” *McKinsey on Government. Special Issue: Defence*, March 2010.

McCarthy, I., and C. Tsinopoulos. “Strategies for Agility: An Evolutionary and Configurational Approach.” *Integrated Manufacturing Systems* 14, no. 2 (2003): 103–113.

“McKinsey on Government. Special Issue: Defence”, March 2010. http://www.mckinsey.com/~/media/McKinsey/dotcom/client_service/Public%20Sector/PDFS/McK%20on%20Govt/Full%20reports/TG_MoG_Issue5_final.ashx.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva P. Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, et al. “Quantitative Analysis of Culture Using Millions of Digitized Books.” *Science* (December 16, 2010). http://www.sciencemag.org/content/early/2010/12/15/science.1199644.abstract.

Ministerie van Defensie. Verkenningen Houvast Voor De Krijgsmacht Van De Toekomst, 2010.

Moffett, S., K. Anderson-Gillespie, and R. McAdam. “Benchmarking and Performance Measurement: a Statistical Analysis.” *Benchmarking: An International Journal* 15, no. 4 (2008): 368–381.

Moriarty, J. P. “A Theory of Benchmarking.” *Benchmarking: An International Journal* 18, no. 4 (2011): 7–7.

Nandi, S., and D. Banwet. "Benchmarking for World Class Manufacturing–concept, Framework and Applications." *Productivity* 41, no. 2 (2000): 189–200.

"NATO - Opinion: NATO – Value for Security'' - Speech by NATO Secretary General Anders Fogh Rasmussen in Bratislava, Slovakia, May 19, 2011, n.d. http://www.nato.int/cps/en/SID-E7F3BCAD-71365729/natolive/opinions_74522.htm.

"NATO - Topic: Defence Planning Process", n.d. http://www.nato.int/cps/en/SID-9EEC64AE-DEFCAC5F/natolive/topics_49202.htm.

"NATO Research & Technology Organisation", n.d. http://www.rta.nato.int/.

"Open Standards Benchmarking Assessments - APQC", n.d. http://www.apqc.org/open-standards-benchmarking-assessments.

Organisation for Economic Co-operation and Development. *Peer Review: An OECD Tool for Co-operation and Change*. OECD, 2003.

Papaioannou, T., H. Rush, and J. Bessant. "Benchmarking as a Policy-making Tool: From the Private to the Public Sector." *Science and Public Policy* 33, no. 2 (2006): 91–102.

"Parameters Advanced Defence Education", n.d. https://spreadsheets.google.com/spreadsheet/ccc?key=0AljscQaoijARdGVDRkpkbFBVLW80bGhsNFZ6clFxdkE&hl=en_US#gid=0.

Raa, Thijs Ten. The Economics of Benchmarking: Measuring Performance for Competitive Advantage. Palgrave Macmillan, 2009.

Rittel, Horst W. J., and Melvin M. Webber. "Dilemmas in a General Theory of Planning." *Policy Sciences* 4, no. 2 (June 1973): 155–169.

Room, Graham. "Policy Benchmarking in the European Union." *Policy Studies* 26, no. 2 (June 2005): 117–132.

Rose, Richard. "Ten Steps in Learning Lessons from Abroad." Working Paper, n.d. http://cadmus.eui.eu/handle/1814/1763.

Seymour, W. A. A History of the Ordnance Survey. Dawson, 1980.

Sommerville, James, and Hamish W. Robertson. "A Scorecard Approach to Benchmarking for Total Quality Construction." *International Journal of Quality & Reliability Management* 17, no. 4/5 (2000): 453–466.

De Spiegeleire, Stephan. Defence Planning: What's Wrong With It and How to Fix It, Forthcoming.

De Spiegeleire, Stephan, Paul van Hooft, Charles Culpepper, and Rene Willems. *Closing the Loop. Towards Strategic Defence Management*. The Hague: The Hague Centre for Strategic Studies, 2009.

De Spiegeleire, Stephan, and Miguette Jadoul. *Learning to Learn Validating the TNO Defence Benchmark Planning Guide*. The Hague: TNO, November 2007.

De Spiegeleire, Stephan. *Towards a Benchmarking Methodology for Defence*. The Hague: TNO, 2006.

Talluri, Srinivas, and Joseph Sarkis. "A Computational Geometry Approach for Benchmarking." *International Journal of Operations & Production Management* 21, no. 1/2 (2001): 210–223.

"The Technical Cooperation Program (TTCP)", n.d. http://www.dtic.mil/ttcp/.

Tillema, Sandra. "Public Sector Benchmarking and Performance Improvement: What Is the Link and Can It Be Improved?" *Public Money & Management* 30, no. 1 (January 2010): 69–75.

———. “Public Sector Organizations’ Use of Benchmarking Information for Performance Improvement.” *Public Performance & Management Review* 30, no. 4 (June 2007): 496–520.

Triantafillou, Peter. “Benchmarking in the Public Sector: A Critical Conceptual Framework.” *Public Administration* 85, no. 3 (September 1, 2007): 829–846.

Ulusoy, G., and I. Ikiz. “Benchmarking Best Manufacturing Practices: a Study into Four Sectors of Turkish Industry.” *International Journal of Operations & Production Management* 21, no. 7 (2001): 1020–1043.

Venkatramaiah, C. *Textbook of Surveying*. Universities Press, n.d.

Voss, Christopher A., Pär Åhlström, and Kate Blackmon. “Benchmarking and Operational Performance: Some Empirical Results.” *Benchmarking: An International Journal* 4, no. 4 (1997): 273–285.

Watson, Gregory H. Strategic Benchmarking Reloaded with Six Sigma: Improve Your Company’s Performance Using Global Best Practice. John Wiley and Sons, 2007.

Wong, Wai Peng, and Kuan Yew Wong. “A Review on Benchmarking of Supply Chain Performance Measures.” *Benchmarking: An International Journal* 15, no. 1 (2008): 25–51.

Zairi, Mohamed. Benchmarking for Best Practice: Continuous Learning Through Sustainable Innovation. Butterworth-Heinemann, 1998.

———. Effective Benchmarking: Learning from the Best. Springer, 1996.

———. Effective Management of Benchmarking Projects: Practical Guidelines and Examples of Best Practice. Butterworth-Heinemann, 1998.

Zairi, Mohamed, and Paul Léonard. *Practical Benchmarking: The Complete Guide*. Springer, 1994.