Counting what counts: performance and effectiveness in the humanitarian sector

Ben Ramalingam and John Mitchell
with John Borton and Kristin Smart

This study looks at the current status of performance approaches in the humanitarian system, in the context of the evolution and use of performance in other sectors. It summarises experiences and potential benefits for the humanitarian sector, and offers recommendations for next steps.

Key message 1

Humanitarian performance should be defined in terms of both its effects on those affected by crisis and according to core humanitarian principles. A more integrated approach to performance, bringing together different levels, functions and initiatives within the system, could help to overcome many of the perceived failures of humanitarian assistance.

There is no widely accepted definition of humanitarian performance, but the working definition developed for this study includes the collective effects of the interdependent humanitarian system of international, national and local agencies, working to save lives, alleviate suffering and maintain human dignity during and in the aftermath of conflicts and disasters. Performance and quality approaches are currently highly fragmented, both conceptually and in practice. A more integrated approach to humanitarian performance would be useful at all levels of the humanitarian system. If conceived and implemented effectively, such an approach could facilitate a stronger, more evidence-based, understanding of the progress made to date and gaps remaining in humanitarian performance.

Key message 2

The international development sector has adopted a system-wide performance agenda in the form of the Millennium Development Goals and the associated aid-effectiveness agenda centred on the 2005 Paris Declaration. At present, there is no equivalent scale or unity of approach within the humanitarian sector.

Performance approaches widely used in the private, public and third sectors include those based on several different traditions in performance management. Results-based management, used widely in many development organisations, focuses on setting goals and objectives, against which performance is then measured. Quality management emphasises above all the perspective of the 'customer' or recipient, and financial audit as a third key tradition. Mixed or balanced approaches, or 'multi-dimensional frameworks', have been used more recently as a way of adapting and integrating ideas from disparate disciplines. The approach adopted in the development sector, around the Millennium Development Goals (MDGs), involves different agencies and countries harmonising operating procedures, emphasis on country ownership, and explicit measurement of performance. Although their achievement has proved uneven and problematic, the MDGs provide a framework for a collective endeavour that does not exist in the humanitarian sector.

Key message 3

The humanitarian system is characterised by a high degree of interdependence. The increasing scale of emergencies, coupled with rising expectations and comparatively limited response capacities, means that even the largest organisation cannot launch an effective response on its own. Despite this, policy apparatus for reflecting on and improving collective performance within the humanitarian sector remains limited.

While there are some cross-organisational efforts to standardise aspects of performance management, there is little consistent or collective working across the whole humanitarian sector. Most performance-related initiatives are still taking place at the level of individual projects or programmes. There are still no baselines, no agreed definitions of performance and an absence of any kind of mechanism able to track performance. The system is still reliant for information on performance from a mixture of different sources, including separate reports, disjointed research and the (very) occasional joint evaluation. There are several approaches to performance management shared by different 'clubs' of organisations, such as NGOs, donors, the Red Cross Movement or UN organisations. There is none that involves all

humanitarian organisations. A system-wide performance framework would have several potential benefits, including the generation of shared indicators of performance which could be compared across different emergencies and over time.

Key message 4

There are many existing projects, initiatives and approaches for monitoring and reporting on performance within the humanitarian system. However, these have been established for a range of different purposes and different focuses and address aspects of performance. Many operate in parallel, and some overlap. Most efforts do not involve regular collection and analysis of data. Those that do are often fragmented in their approach.

Existing performance-related approaches and initiatives can be seen to focus on five different aspects of the system – (1) context, needs and inputs at the global and local levels (2) projects and programmes (3) whole operations (4) organisations (5) system-wide performance. They also address different aspects of performance, often in parallel or with overlap. Many initiatives do not systematically gather data against their performance criteria, making it hard to track progress against their specific goals for improving the system. There is an increase in initiatives that do track data – especially against indicators such as mortality and nutrition, and the last few years have seen steps towards improved gathering and sharing of information and data across such initiatives. The problem now is not so much lack of information on humanitarian assistance but *fragmentation* between monitoring and measurement systems which have often been set up for contrasting reasons. These are based on different methodologies and categorised according to different definitions without common standards. There is clearly scope for the range of available approaches and methods to be usefully synthesised, both at the conceptual level and also in terms of the data available. Such a synthesis could help to foster an effective and coherent analysis of humanitarian responses.

Key message 5

Despite the considerable number of systems gathering data, and the range of initiatives attempting to address performance issues, most of these initiatives tend not to use the available data to improve performance. There is often no connection between data collection and the use or application of those data for systematic reflection and learning.

Much assessment and monitoring effort in the sector focuses on what should be done, often in accordance with stated objectives as laid out in proposals, and stop short of reflecting on what was actually done, and how well. As a result, existing performance mechanisms – such reporting to donors – are generally not used to their full potential and are instead becoming a disconnected administrative chore. Systems established to track key indicators such as mortality and nutrition have been developed outside organisations. These tend not to include any way of establishing attribution of humanitarian performance to the changes they monitor, and tend not to be used within agencies to inform their analysis. Even where performance systems are reasonably well-developed – for example, results-based management systems in multilateral humanitarian agencies – they are seldom used as the basis for management decisions and resource allocation.

Key message 6

While evaluation plays a central role in current humanitarian performance, and has many benefits, it may not be adequate for the longitudinal monitoring of trends in performance. Evaluation and results-based performance management are often regarded as alternatives but they have different and complementary functions.

Evaluation can be tailored to the needs of the agency concerned, and can generate learning and accountability in some form even where monitoring systems are weak or absent. Evaluation also allows detailed exploration and explanation of what happened and why, using both quantitative and qualitative approaches ideal for understanding context and generating recommendations. The likelihood of more

joint, system-wide evaluations of selected operations is a welcome development and will provide an important source of information for assessments of the overall performance of the humanitarian system. However, the specific and highly customised nature of evaluations makes it difficult to compare results directly between them. In contrast, performance monitoring determines whether specific results were achieved, using mostly quantitative indicators, and enables comparability across a wider scale and range of activities, and over time. Any integrated performance framework needs to intelligently combine performance monitoring with evaluation in order to fulfil its potential.

Key message 7

While the humanitarian system already has many of the elements necessary for a comprehensive model of performance, there are several key areas with little or no information or inadequate methods of inclusion and analysis.

Four outstanding gaps in humanitarian performance management are identified in this study. First, the views and opinions of key stakeholders, especially beneficiaries and affected populations, are underutilised. Other stakeholders, including partners, Western donor publics and national governments, also appear to be routinely left out of performance considerations. The relative lack of mechanisms within performance approaches for seeking opinions at beneficiary and community levels bears out the frequent criticism of the humanitarian system that it does not make sufficient effort to gather beneficiary views, and pays little regard to such views when they are available. Second, there appears to considerable scope for improving analysis of how contexts affects and can constrain performance. This would require some methodological development and testing, but such analysis has the potential to inform operational planning as well as subsequent performance assessments. Third, there is scope for more consideration of organisational capacities, especially in terms of capacity strengthening, partnerships and innovation (the last of which is looked at in an accompanying study in this year's Review of Humanitarian Action). The fourth and final gap in performance efforts is the lack of focus on impact and outcomes, although there is some evidence that this is beginning to change (again, the focus of an accompanying study).

> **Key message 8**
>
> **In moving towards a more coherent and consistent approach to performance management in the humanitarian sector, a useful first step would be to integrate and balance the existing variety of concepts and methods. This would also reflect the fundamental interdependence of the modern humanitarian system.**

A balanced, comprehensive and coherent framework for humanitarian performance could be discussed and agreed across agencies. Such a framework could draw on and adapt aspects from a range of different approaches to performance and quality management. Existing efforts and initiatives could then be mapped and collated against the framework. This approach could also enable the identification and clarification of key issues and gaps within the humanitarian system. It could for example be used to improve harmonisation of processes for performance reporting, thereby reducing the administrative burden of existing reporting systems. It could also be used to prioritise gaps to be filled by sustained, sector-wide efforts in research and development, such as how to incorporate beneficiary perspectives, context analysis, capacity strengthening and impact assessment. With the agreement of common case-study or pilot countries within current and planned projects, work could be synthesised across initiatives to generate a series of integrated and comprehensive views of performance in the selected settings.

> **Key message 9**
>
> **The ongoing and regular synthesis of system-wide performance would be a major step further towards a more relevant, strategic and credible international humanitarian system. A regular synthesis could enable preliminary assessment of the 'state of the system', and set a baseline for establishing collective goals and assessing progress against them.**

There have been some landmark reviews of the overall humanitarian system, such as the 2005 *Humanitarian Response Review*, and there are key ongoing series such as the *World Disasters Report*. However, these reviews focus on particular elements of humanitarian performance, or vary their themes from year to year, and so do not provide consistent, regular analysis of the same issues. A regular synthesis of the

whole humanitarian system could begin with a pilot assessment looking at performance using existing data, supplemented with tools such as interviews and surveys to cover selected data gaps. Eventually, such an approach could be used to set shared, system-wide goals and to track progress towards them. In this way, performance could provide an overarching framework for understanding the vast array of humanitarian initiatives, approaches and techniques. The challenge is to reframe all these initiatives in terms of their separate but complementary contributions to the effectiveness of humanitarian aid. Doing so would help to move considerations of humanitarian performance beyond narrow snapshots to a more longer-term, strategic perspective. In doing so it is important to be aware of the problems associated with the performance agenda, and to carefully navigate them whilst harnessing the benefits. In their ongoing effort to improve performance, actors in the sector would do well to remember the words of Albert Einstein: 'not everything that can be counted counts, and not everything that counts can be counted'.

1.1   Introducing performance

1.1.1   Aims and scope of this study

At the heart of ALNAP's mission, as laid out in its five-year strategy, is the notion of 'fostering a culture of active learning and accountability to improve humanitarian performance'. In 2006, following the publication of the reports of the Tsunami Evaluation Coalition (TEC), ALNAP launched what would become the Humanitarian Performance Project (HPP), an initiative to reflect on, review and track changes in system-wide performance on a regular basis. The motivation was the strong sense that the system has so far come together to review its overall performance only after major crises – and that the ensuing analysis was very much focused on those specific emergencies, such as Goma or the Indian Ocean tsunami.

The three years of the HPP, with multiple iterations and different paths explored, has shown that a more comprehensive and integrated understanding of approaches to humanitarian performance has the potential to inform not just this project but also the range of performance-related initiatives within and across humanitarian organisations. This first study of the three comprising ALNAP's 8th Review of Humanitarian Action has two overarching aims.

1   To present an overview of approaches to performance management, illustrated with examples from different schools of thought and areas of practice (Section 1.1). This includes private- and public-sector experience in developed countries, but also presents approaches in international aid such as the Millennium Development Goals, the Paris Declaration on Aid Effectiveness and results-based management (RBM). System-wide performance approaches are also examined, in terms of the value they add to organisation-specific approaches and some of the challenges they face.

2   To relate these ideas to the international humanitarian system. Following a suggested definition of humanitarian performance, based on shared principles and objectives, Section 1.2 reviews the application of performance approaches to humanitarian work in different contexts and in relation to different aspects of the system (context, needs and inputs; projects and programmes, operations, organisations and system-wide).

Given the sheer number of initiatives 'out there', this review does not attempt to be comprehensive. A more detailed compilation of the different systems and approaches will be published on the ALNAP website page for the Humanitarian Performance Project to accompany this study (Borton, 2008a). In this chapter we want to focus on providing an illustrative overview, to highlight the key issues, gaps and challenges facing approaches to performance in the humanitarian sector. The final section of this chapter (Section 1.3) presents an initial conceptualisation of an integrated, comprehensive approach to humanitarian performance. Building on this, a series of recommendations are offered to improve and strengthen performance approaches across and within humanitarian organisations.

1.1.2   The evolution of performance approaches in the private sector

Although the term 'performance' is widely used, it tends to be defined only indirectly and according to context. The literature on performance has no solid, common theoretical basis, and there is a vast diversity of views on what performance is and what can be done to improve it. Contributions from academics, quasi-academics, think-tanks, consultancies, practitioners, and professionals or managers within different organisations create a dizzying variety of performance-related methodologies and frameworks (Neely et al, 2007). Moreover, performance is frequently presented as an umbrella for a host of other ideas – including effectiveness, productivity, quality, transparency and accountability – each of which leads to yet more frameworks and extensive literatures.

This section (1.1.2) aims to provide context for the exploration of humanitarian performance, by summarising the history of performance in terms of key thinkers, practitioners and legacies. It looks especially at the evolution of performance approaches in the context of industrialisation in the private sector. This is intended to provide a broad overview of how performance has been, and is now, understood and used in modern organisations.

The use of performance data to monitor, evaluate and improve the effectiveness of organisations arguably dates back at least as far as Ancient Egypt, where the work of the pyramid-builders demonstrated 'an ability to envision, organise and manage complex activity dependent on reliable data that pre-dates the modern world' (Lawrie et al., 2005, p. 2). For many of the centuries that followed, across many

different societies, the notion of performance was closely connected with three distinct fields.

1   Military strategy has long aimed to define goals to be achieved in a specific military campaign, and to devise tactics to achieve those goals. Military strategy, and the hierarchy of objectives within it, has had a profound influence on modern management thinking.

2   Public tax collection in many different countries has involved analytical and survey techniques to find out what was owned by citizens (such as land and livestock), and its worth, and then to determine the taxes due. The first such national survey in Britain is recorded in the Domesday Book, after the Old English word meaning 'accounting' or 'reckoning'. This approach is based on the principles of audit: surveys, accurate documentation and reporting, controls for ensuring payments and penalties for shortfalls.

3   Organisations of skilled craftspeople or professionals, from painters or sculptors to doctors or lawyers, have involved senior practitioners working to train and supervise the work of juniors and apprentices. The senior practitioners thereby set standards, review the work of others and order rework and revision as necessary. Craft and professional approaches to performance have contributed significantly to modern quality management.

The intellectual foundation for the industrial conception of performance was laid down by Adam Smith, in *The Wealth of Nations*, comparing factory methods to previous methods for making pins.

'But in the way in which this business is now carried on... it is divided into a number of branches, of which the greater part are likewise peculiar trades... the important business of making a pin is, in this manner, divided into… about eighteen distinct operations... I have seen a small manufactory of this kind where ten men only were employed... they could make among them about... forty-eight thousand pins in a day... But if they had all wrought separately and independently, and without any of them having been educated to this peculiar business, they certainly could not each of them have made twenty, perhaps not one pin in a day; that is… perhaps not the four thousand eight hundredth part of what they are at present capable of performing, in consequence of a proper division and combination of their different operations' (Smith, 1776).

This extract includes rational analysis of the nature of the undertaking, tasks involved, how the tasks are performed, and measurement of efficiency by comparative results. These four elements, although much evolved and adapted, remain at the heart of modern approaches to performance. Smith's ideas and findings were implemented on a large scale by early British industrialists such as Josiah Wedgwood, pioneering owner of Wedgwood China. Many early industrialists and management thinkers in other countries at this time consciously followed the British methods of industrialisation, in much the same way as Japanese firms would later seek to emulate US firms, and then firms globally sought to emulate the resulting success of Japanese firms.

Foundations of modern performance management

Around the turn of the nineteenth century, three men working in different countries – Frederick Taylor in the US, Henri Fayol in France and Max Weber in Germany – separately undertook work now widely seen as the classical basis of managing performance. Their pioneering work contributed to industrial and public sector practices throughout the twentieth century and as we shall see form the basis of many modern approaches.

Frederick Taylor worked in America at the start of the twentieth century. A mechanical engineer by training, he was interested in optimising industrial efficiency in factories. Detailed analysis of factory working methods led Taylor to argue that, through observation, measurement and effective oversight, a 'scientific' approach to efficiency could be established by managers. He based his 'scientific management' approach on five principles.

1   Replace rule-of-thumb work methods with methods based on a scientific study of the tasks.

2   Scientifically select, train, and develop each employee rather than passively leaving them to train themselves.

3   Provide 'detailed instruction and supervision of each worker in the performance of that worker's discrete task'.

4   Divide work nearly equally between managers and workers, so that the managers apply scientific management principles to planning the work, and the workers perform the tasks without interruption.

5 Provide wage incentives for workers with highest levels of output. (Taylor, 1911; Kanigel, 1997).

Scientific management focused on producing the same or greater output with lower costs, and Taylor regarded the role of individual workers as very limited. He argued that performance improvements were possible only through managers enforcing top-down approaches. In contrast, several influential studies in the 1920s examined the effects of social relations, motivation and employee satisfaction on factory productivity. This work provided the platform for the human-relations school, which contrasted with scientific management in its view of workers in terms of psychology and 'fit' with companies, rather than as interchangeable parts. According to one analyst, 'the hallmark of human-relation theories is the primacy given to organizations as human cooperative systems rather than mechanical contraptions' (Burnes, 2004).

In Europe, working at around the same time as Taylor, Henri Fayol was a mining engineer who became managing director of a successful mining corporation, and was credited with turning around the fortunes of the organisation. In retirement he set up the Centre d'Etudes Administratives, a pioneering management education establishment. Fayol identified a number of 'universal principles' applicable to business or public administration. These have aspects in common with Taylor's approach, in that they include the division of work, the use of rewards and sanctions, unity and chain of command, fair remuneration, subordination of individual to organisational interests, fairness, stability, initiative and esprit de corps (Burnes, 2004).

Unlike Taylor, however, Fayol did not present these ideas as an attack on previous management approaches, but as a form of education. Fayol described an 'administrative science' – a consistent set of principles that managers needed to learn in order to help their organisations run properly. Some attempts were made to promote Fayolisme as an alternative to Taylorism, but others – including Fayol himself – saw them as complementary. While Taylor focused on the operational level and on work methods, measurement and simplification to secure efficiencies, Fayol focused on the leadership level, and on principles of organisation and management, Despite their differences, these two influential thinkers had more that united than separated them.

Third, Max Weber, an influential German political scientist, described the 'ideal bureaucracy', drawing many of his ideas from the public administration of Prussia

(Burnes, 2004). The ideal bureaucracy was a public-sector equivalent of the work of Taylor and Fayol, with six essential features: specialisation; hierarchy; written rules and regulations; impartiality; impersonality; and record-keeping. These features were seen as essential to delivering bureaucratic efficiency.

After, Taylor, Fayol and Weber, much of the work on performance was shaped by the efforts of successful business leaders and managers, who demonstrated the value of performance systems for business development. Three are of particular interest here, from the 1910s and 1920s, because they each inspired different approaches and schools of thinking around performance which are still in use today.

1 Henry Ford introduced scientific management principles into the mass production of Ford Model-T automobiles. His success contributed to high regard for the role of science in management efficiency, and to the wide spread of scientific management principles.

2 Walter Shewhart, working at US telephone company, Bell, set out the principles at the heart of quality control, focusing on reducing variation in a manufacturing process to ensure quality of outputs.

3 The Dupont Company developed a performance system to link financial metrics of performance to return on investment. This moved the application of scientific management from shop floor to financial and management accounting approaches.

Each of these proved tremendously influential on subsequent approaches. Scientific management principles became central to 'management by objectives' and subsequently results-based management. Quality control processes shaped the subsequent quality-management movement. And modern systems of audit and financial management owe a debt to the work of Dupont. We will look at each of these in turn below.

Management by objectives (MBO) and results-based management (RBM)

Management by objectives (MBO) is the forerunner to results-based management, as widely used today in the public sector. In the 1950s, influential management thinker Peter Drucker published *The Practice of Management*, containing the principles of MBO (Drucker, 1955; Waring, 1994). This approach aims to create empowered

employees, clear about their roles and responsibilities, and therefore able to contribute to the achievement of organisational as well as personal goals.

MBO involves five key steps, as follows.

1 Establish a set of top-level strategic goals and objectives.

2 Create a cascade of organisational goals and objectives supported by lower-level objectives and action plans.

3 Develop an organisational role and mission statement, as well as specific objectives and action plans for each member, often involving participative decision making.

4 Establish key results and/or performance standards for each objective.

5 Periodically assess the status or outcome of the goals and objectives.

Key features of MBO include:

- motivation – involving employees in the whole process of goal-setting, and increasing employee empowerment, increases job satisfaction and commitment

- better communication and coordination – frequent reviews and interactions between superiors and subordinates help to maintain harmonious relationships within the enterprise, and also solve many problems

- clarity of goals – through the use of SMART objectives (specific, measurable, achievable, relevant, and time-bound)

- objectives set at both individual and collective levels, and in all domains of the organisation, from research to production to finance information systems

Quality management

W. Edwards Deming first applied what would become Total Quality Management (TQM) to the production of munitions during World War II in America, and developed his ideas in the 1950s. They were adopted enthusiastically in Japan, where quality improvement was made a central element of rebuilding the post-war

economy. Deming, his forerunner Walter Shewhart, and their ideas played an important role in the subsequent success of Japanese products, made with a focus on overall processes and customer satisfaction as a critical measure of quality. The success of Japanese products in American markets then led US businesses to explore quality approaches, borrowing back from Japan as well as using newer US approaches (Kearney et al., 1999).

Within the principles of quality management, the first and key idea is that the customer is the ultimate arbiter of product quality. As a result of this customer focus, quality should be built into the product early in the production process rather than being added on at the end; many products and services go through design and production, with no quality check except through later customer complaints. TQM opposes quality control through inspections and complaints, because this shifts responsibility away from initial designers and producers (ibid.).

The concept of product quality within TQM is based on reducing variability, using process-control charts to track deviation from an optimal level. Quality requires continuous improvement of inputs and processes, being represented by a 'delighted customer': if customer expectations rise, so too must quality. Quality also results from people working within systems, rather than from individual efforts. Equally, when quality slips, it is the system rather than the person that should take the blame. A well-performing system should lead all workers to perform well, and so within TQM, merit-based pay is seen as a 'lottery', as is managing by objectives. Quality within TQM requires strong management commitment to high standards of performance, and strong worker participation, enabled by a process of working without fear of mistakes being punished, and also of working without barriers between hierarchies and functions.

There is some confusion about the difference between TQM, which has had several manifestations, and other quality approaches. Standards of the European Foundation for Quality Management (EFQM) are an adaptation of TQM processes. They represent a holistic, systemic approach to management, based on a set of generic principles evolved from Shewhart's pre-war practices of statistical process control (Macpherson, 1999).

By contrast, the International Organization for Standardization (ISO) approach involves an internationally recognised standard analogous to a warrant of fitness applied to an organisation's procedures and systems. As with TQM, it was originally manufacturers,

especially those selling in international markets, which sought ISO certification, followed by suppliers in the raw materials, transportation and distribution industries. More recently, service providers such as legal firms and insurance companies have sought certification. This has also spread to government and NGOs seeking to standardise the quality of their products and processes (Macpherson, 1999).

Quality management is seen by its proponents as very different from management by objectives, and Deming, the developer of TQM, argued that management by objectives works against the culture of support at the heart of total quality. Deming went so far as to define TQM in terms of the elimination of MBO approaches. In his view, 'a focus on outcomes must be abolished, and leadership put in its place' (Kearney et al., 1999). In particular, he argued that an MBO approach leads to a component-based view of performance, and that this could lead to making specific teams or units the focus of blame. Instead, leadership is needed to provide a systemic view, providing a perspective on the interdependence of different functions within the organisation, and making appropriate suggestions. This was seen to be in contrast with MBO approaches, wherein managers using only objectives might locate blame where it was not justified.

Financial approaches to performance management

The third performance-management movement that developed in the twentieth century was the use of accounting and financial ratios as a means of assessing business performance. Developed from the main sources of financial information of balance sheets, income statements and cash-flow statements, financial ratios including return on investment and return on capital employed were used to generate regular management reports about business performance. Such mechanisms provide an objective and comparable means of motivation and control within organisations (Otley, 2002). These approaches usually involve increasing the overall value of the organisation, in terms of what can be attributable to shareholders. This is usually in terms of 'increasing shareholder value'. There are a number of different levels of analysis that this involves.

The first level of analysis focuses on capturing controllable aspects of performance in earnings and balance sheet measures. This involves the use of financial indicators to represent the underlying activities being managed. The financial tools provide common language for very different activities to be analysed in terms of measures such as sales revenue, costs and profitability. This usually means dividing the

organisation into profit centres – which are responsible for generating sales revenue and managing costs, and cost centres, which are responsible only for cost control.

Multi-dimensional approaches to performance management

The 1980s and 1990s saw the emergence of mixed approaches, or 'multi-dimensional' frameworks, in reaction to the partial approaches to performance that resulted from each of the MBO, TQM and financial movements. The mixed approaches emphasise the combination of strategic, financial, internal process and customer/supplier perspectives into a single balanced framework. Many of these were developed with a sense that pre-existing frameworks said 'nothing about the factors that actually help grow market share and profits – things like… innovation, R&D effectiveness, quality, employee development' (cited in Neely, 2007). A recent review notes that it is only relatively lately that ideas from disparate disciplines have begun to converge in recognition of the need for integration into a multi-disciplinary approach (Neely et al., 2008). In effect, what these approaches achieved was to combine a range of different concerns, providing a framework for prompting managers to think through the range of issues relevant to their organisation's performance.

Although balanced models were in use in some form in large companies such as General Electric as far back as the 1950s, there is now a range of mixed methods and approaches that dominate the performance landscape (I&DEA, 2006; Neely, 2007; Thiel and Leeuw, 2002). One particular model, the 'Balanced Scorecard' (BSC) proposed by Kaplan and Norton in 1992 (Kaplan and Norton, 1992) has been particularly influential, and there is evidence showing that by 2001 it had been adopted by 44 per cent of firms worldwide (Neely et al., 2007). As the balanced approaches have taken off, there is also a growing awareness of the need to blend approaches in order to help to address gaps in each framework. For example, 2GC, a leading performance consultancy, has combined the Balanced Scorecard and EFQM approaches to good effect (Lawrie et al., 2005). Box 1.1 shows the range of performance indicators typically used in the private sector today.

A number of mechanisms exist to model performance, and increasingly, these combine different approaches. Perhaps the most widely used mechanism is the balanced scorecard and strategy map, which as illustrated in figure 1.1 below, can be used to map the full range of different business objectives and determine the weighting given to each objective, as well as to reflect the assumed causal relationships between different objectives.

Box 1.1   Indicators of performance used in the private sector

One recent study reviewed ten distinct performance approaches – from EFQM to the Balanced Scorecard – and found considerable similarity between approaches in terms of the performance dimensions covered (Hudson et al., 2007). The list below illustrates the findings of the study, which grouped performance into six general dimensions.

Finance	Time	Flexibility
Cash flow	Lead time	Manufacturing effectiveness
Future growth	Process throughput	Volume flexibility
Market share	Process time	New product introduction
Overhead cost reduction	Productivity	IT systems
Inventory performance	Cycle time	Future growth
Cost control	Delivery speed	Product innovation
Sales/revenues	Labour efficiency	
Profitability	Resource utilisation	
Efficiency		
Product cost reduction		

Quality	Human resources	Customer satisfaction
Product performance	Employee relations	Market share
Delivery reliability	Employee involvement	Service
Waste	Employee skills	Image
Dependability	Quality of work life	Integration with clients
		Competitiveness
		Innovation

These six dimensions can be seen to cover all aspects of organisations: operating performance (through the dimensions of quality, time and flexibility; financial results; how the company is perceived externally (through its customers); and aspects of the working environment (through the human-resource dimension). A firm might use any combination of these as part of a suite of key performance indicators (KPIs) – 'a set of measures focussing on those aspects of organisational performance that are most critical for the current and future success of the organisation' (Parmenter, 2007).

Figure 1.1 The balanced scorecard and strategy map

FINANCIAL PERSPECTIVE: THE DRIVERS OF SHAREHOLDER VALUE

Improve shareholder value

Shareholder value ROCE

Revenue growth strategy   **Productivity strategy**

| **Value from new products and customers** | **Increase customer value** | **Improve cost structure** | **Improve asset utilisation** |

New revenue sources   **Customer profitability**   **Cost per unit**   **Asset utilisation**

CUSTOMER PERSPECTIVE: THE DIFFERENTIATING VALUE PROPOSITION

Customer acquisition   **Customer relation**   **Market share**

Product leadership

Complete customer solutions

Customer value proposition   Value for money

Product service attributes   Relationship   Image

Price   Quality   Time   Function   Service   Relations   Brand

Customer satisfaction

INTERNAL PERSPECTIVE: HOW VALUE IS CREATED AND SUSTAINED

| **"Innovate"** (processes that create new products and services) | **"Increase customer value"** (customer management process) | **"Achieve operational excellence"** (operations and logistics processes) | **"Be a good neighbour"** (regulatory and environmental processes) |

LEARNING AND GROWTH PERSPECTIVE: ROLE FOR INTANGIBLE ASSETS – PEOPLE, SYSTEMS, CLIMATE AND CULTURE

A motivated and prepared workforce

Strategic competencies   Strategic technologies   Culture and climate for action

Source Kaplan/Norton, 2000

1.1.3 Performance approaches in the public sector

'The government are very keen on amassing statistics. They collect them, add them, raise them to the nth power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first instance from the village watchman, who just puts down what he damn well pleases' (Josiah Stamp, Head of the Inland Revenue Department of the UK, 1896–1919).

Performance is not a new area of interest in the public sector, and large-scale organisational management is arguably almost exclusively a public-sector activity for much of history (Lawrie et al., 2005), until the early industrial era. As in the private sector, the nineteenth century saw the introduction of rational models which profoundly influenced all walks of life. For example, the tremendous increase in scientific knowledge radically altered and expanded the practice of medicine, with a shift from the master–apprentice model of teaching clinical medicine to a more democratic system of medical schools. Concern for the quality of services led to the establishment of public and professional bodies to govern standards for medical training and practice. The use of evidence became central, as in Florence Nightingale's use of statistical methods to demonstrate how better hygiene can reduce war-hospital mortality rates.

After the Second World War, performance and quality management schemes were introduced across developed-country public sectors, with some attempts at use of targets and performance indicators. Early performance measures were concerned with 'input' aspects, mainly focusing on financial resources. Different methods were trialled: cost-benefit analysis in the 1960s, output-based budgeting in the 1960s and management by objectives (MBO) in the 1960s and 1970s. Most of these initiatives, however, were regarded as experimental and largely one-off efforts (UNDESA, 2003).

It wasn't until the 1980s and 1990s that performance concerns became central to public-sector practice (Hailey and Sorgenfrei, 2004). The New Public Management (NPM) movement arose in response to a number of pressures, especially:

- economic pressures, including budget deficits, structural problems, growing competitiveness and globalisation

- political and social pressures, including a lack of public confidence in

government, growing demands for better and more responsive services, and better accountability for achieving results with taxpayers' money.

At its heart, NPM has consisted of a wave of deliberate reforms to structures and processes of public organisations, with the intention of getting them to perform better (Pollitt and Bouckaert, 2004). The principal element of these reforms was a shift in management focus from inputs and processes to outputs and outcomes, with a concurrent shift towards more assessment and measurement. This was manifested by performance indicators and standards to ensure that government activities achieved desired results (Pollit, 2001). Underlying these changes was a broader political shift towards much wider use of market or market-like mechanisms for the delivery of public services, either directly or through the rubric of 'choice' for clients of state services. A study of the experiences of ten OECD countries in introducing NPM approaches showed that performance management was a key plank in the efforts of all ten (Meier, 2003).

Performance-oriented slogans such as 'Reinventing government', 'Doing more with less' and 'Demonstrating value for money' summarise the approach of these reforms. While each country has implemented reforms according to particular political, institutional and legislative contexts, there have also been a number of common aspects across countries (OECD, 2000). These include:
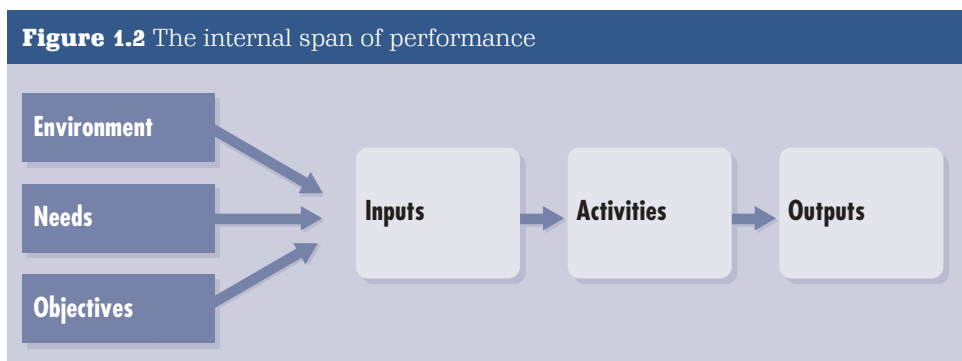
- focus on performance issues, such as efficiency, effectiveness, quality and value of services

- devolution of management authority and responsibility

- orientation to customer needs and preferences

- participation by stakeholders

- reform of budget processes and financial management systems

- application of modern management practices.

NPM has involved the growing use of performance approaches, with thousands of performance indicators and frameworks introduced to allow various governments to control costs, increase accountability, demonstrate performance and improve

services. Some of these include TQM programmes, Charter Marks, the Investor in People scheme, and the Balanced Scorecard.

Two important concepts for understanding performance approaches in the public sector are depth and span (Bouckaert and Halligan, 2008). The *depth* of performance relates to the different levels at which performance might be assessed. This includes the micro-level of projects and programmes; the meso-level of particular policy fields, specific sectors, or specific service or delivery chains; and finally the macro-level of entire governance systems or cross-sector-wide approaches.

Also important is the *span of performance* in the public sector. This is typically illustrated with reference to some external need and environmental trigger, to which the public sector responds through setting objectives, mobilising inputs that are processed in activities, which result in outputs. This is typically represented as a linear process, or chain, as shown below.

Figure 1.2 The internal span of performance

Internal performance can focus on any of these areas individually, or in relation to each other. For example,

- assessing inputs in relation to needs can identify the *relevance* and *appropriateness* of an intervention

- assessing outputs in relation to inputs gives an indicator of *efficiency*.

Such analyses tend to assume and rely on a direct link between inputs, activities and outputs, and can be taken as a linear, machine-based approach akin to scientific

management. However, in the public sector, this chain is full of disconnections, disruptions and disjunctions, all of which are potential sources of inefficiency and performance problems.

As highlighted earlier, outputs are no longer seen as an end in themselves in public sector performance. Outcomes or impacts are those changes which take place as a result of outputs, and they are influenced by many external factors, often beyond the control of any single organisation.

If outcomes and impacts can be assessed, a comparison with outputs can give an indication of the *effectiveness* of a programme or project. It has been argued by leading public sector performance analysts that effectiveness, defined in these terms, should be the primary dimension of performance, while efficiency and other internally-focused criteria, are secondary dimensions. The ideal model is presented in Figure 1.3 below.

However, this is not unproblematic. This has been referred to as a major disconnect – a 'Grand Canyon' – that exists "between outputs on the one side, and a disrupted and distant, almost unreachable, but visible sequence of... [outcomes and impacts] on the other side..." (Bouckhaert and Halligan, 2008, p. 17). Addressing this requires carefully defining outcomes and impacts, and determining the underlying programme logic which will enable them to be achieved.

It may be the case that even assessing impacts may not be sufficient. Arguably, the ultimate ambition of performance in the public sector is to strengthen learning and accountability, and thereby establish trust among a variety of stakeholders. This is embodied in approaches such as performance-based budgeting, where allocation of
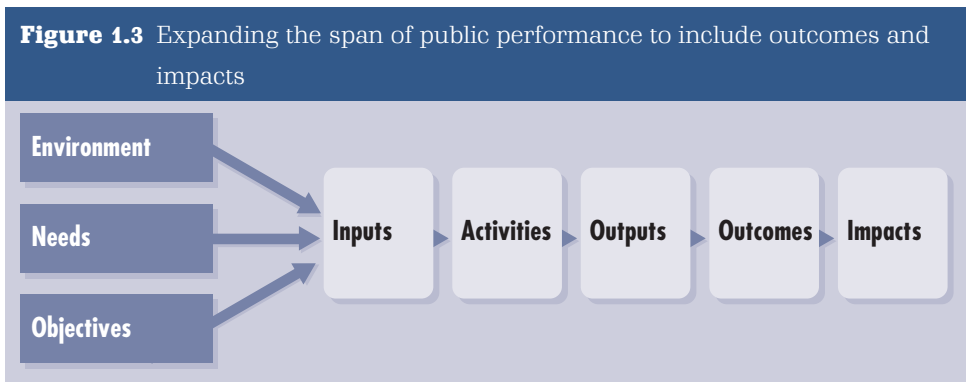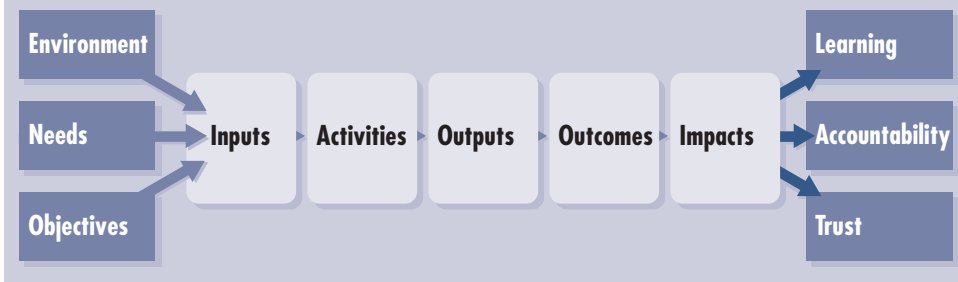
Figure 1.3 Expanding the span of public performance to include outcomes and impacts

Environment

Needs

Objectives

Inputs → Activities → Outputs → Outcomes → Impacts

Figure 1.4 Expanded span of public sector performance to include learning, accountability and trust

resources according to performance is underpinned by the idea that some programmes can be trusted more than others. However, there are increasing numbers of arguments against the idea that improving effectiveness necessarily increases trust. Figure 1.4 illustrates this.

This is a second 'Grand Canyon' of public sector performance (Bouckhaert and Halligan, 2008), and can only be bridged through the engagement of all stakeholders, and serious and honest reflections on the role and purpose of the public sector. Paradoxically, some analysts suggest that the approaches to performance that have been implemented in the past decades may have actually reduced trust in, and accountability of, public services.

The above discussion is pertinent to any discussion of performance because different depths and different spans of performance are required for different purposes. Combining span and depth means matching environmental triggers, needs and objectives, with inputs, activities, outputs, outcomes and impacts at different depths. This can be done for a variety of strategic purposes, which contribute to a greater level of trust within and across the public sector.

In general, performance approaches are introduced with the twin purpose of management improvement (learning and innovation) and performance reporting (accountability and transparency) (Meier, 2003; de Bruijn, 2007). The literature indicates a general tendency to emphasise performance for accountability and external reporting over performance for management, which echoes the situation of evaluation in the aid sector. This can lead to distortions in how performance influences activities. The common shorthand for this effect in the private sector,

'what gets measured gets done' can be applied in public-sector contexts as 'what gets targeted and reported gets done' (I&DEA, 2006).

The two aims of management and accountability may be conflicting and entail somewhat different approaches and systems. Performance for management improvement focuses on gathering and using performance information to support learning and decision-making, for example by feeding back results to make routine adjustments to improve programmes, or to allocate more resources to the best-performing programmes. This model of 'managing-for-results' has proved challenging in many OECD governments, as performance information for management decision-making requires the performance system to be integrated throughout key systems and processes, and because performance-based resource allocation is not straightforward and cannot be rigidly applied. While it may appear to make sense to reward organisations and programmes that perform best, punishing weaker performers may not always be feasible or desirable.

The second aim of accountability for results, and to ensure transparent reporting of those results, when performance information is used for reporting to external audiences, is referred to as 'accountability for results'. Such reporting can be mandated by legislation or executive orders, and can be useful in the competition for funds by convincing a sceptical public or legislature that an agency's programmes produce significant results and provide 'value for money'. Annual performance reports may be directed to a variety of stakeholders, including ministers, parliament, audit and oversight agencies, clients and the general public.

Overall, these public-sector performance approaches appear to draw mainly on private-sector approaches. Although many elements of the NPM movement were presented as new initiatives, they were based – especially in the early stages – on the principles of management by objectives.

While many performance frameworks can be applied in contexts beyond the business world, the different approaches may not fit together well, if at all. For example, frameworks such as MBO, Total Quality Management, Balanced Scorecard and EFQM were all taken up within public services in the 1980s and early 1990s, often by different units with different aims. The conceptual tensions between these approaches have frequently been overlooked, and different frameworks have often been applied concurrently, creating both confusion and 'initiative overload'. Despite

this blurring of different approaches, the NPM movement – with its focus on budget allocations, results, and control measures – indicates the profound influence of scientific management, MBO, and audit-and-control approaches (Kearney et al., 1999; Bourgon, 2008).

New Public Management in international development

Public-sector bodies concerned with international aid and development have been grappling with performance issues for some time. Because of the close relationship between humanitarian and development work, this effort is included here. For much of its history, development aid has been dictated by agendas and interests arising outside developing countries, and in ways that were largely unresponsive to national agendas. Much aid flow has focused on funding inputs and activities, rather than on achieving tangible development goals. The rise of New Public Management coincided with 'reform and reformulation' in development aid (World Bank, 2008), with greater awareness of the need for improved measures for effective, results-oriented aid efforts. This was articulated within donor agencies and aid-receiving sectors in developing countries (Boxes 1.2 and 1.3).

Box 1.2	New Public Management at DFID

The UK Department for International Development (DFID) has identified a series of performance indicators and targets as part of the Public Service Agreements (PSAs) which DFID agrees with the Treasury. The PSA is used as a tool to manage and improve performance, providing a high-level framework against which policy decisions and financial commitments can be assessed, and successes and under-performance measured. DFID's management processes are set up so that at each level – form division to departments, teams and individuals – targets can be set and measured that relate directly to the achievement of the PSA objectives. Progress towards each target is tracked continuously and formally reported twice a year in DFID's Autumn Performance Report and the Annual Report). The PSAs for 2003–06 and 2005–08 are based on the Millennium Development Goals (MDGs) and so the targets and indicators selected relate to the MDGs rather than the agreed objectives of humanitarian action – the saving of life, alleviation of suffering and maintenance of human dignity.

Box 1.3	New Public Management in the Ghana water system

In little more than a decade, Ghana has transformed the structure and strategy of its rural water-supply sector. By 2000, district assemblies and communities played a significant role in planning supplies. The new policy and structure has attracted extra funds, and work is accelerating. This reform process started with an extended dialogue with the major stakeholders in the sector, from which a new policy on rural water and sanitation was developed. The policy was then implemented in several large pilot projects, supported by a number of external agencies, and finally the lessons from those projects were incorporated into the national programme itself.

The success of this approach is attributed to the contracting of national and international NGOs to build the capacity of local-level NGOs and community organisations. The Community Water Supply Agency (CWSA) was created as a facilitating agency rather than an implementer. As a semi-autonomous public-sector agency, CWSA signs an annual performance contract with the State Enterprise Commission. It is committed to staying efficient and small, below 200 staff, and is highly decentralised to its ten regional offices.

These changes have also had a major influence on the overall policy context in which development aid takes place, notably through the Millennium Development Campaign, within which 189 UN member countries agreed in 2000 to common global aims for poverty reduction. These were reflected in the eight Millennium Development Goals (Box 1.4 overleaf), with 18 targets and 48 associated performance indicators, to be achieved by 2015.

It was established that the movement towards the goals would be tracked on a regular basis at country, regional and global levels. As part of this effort, there was a growing recognition of the need to identify the programmes, mechanisms and strategies that can best realise the MDGs. This effort, collectively labelled the aid-effectiveness agenda, has involved a series of high-level discussions culminating in the 2005 Paris Declaration.

> '[W]e the ministers of developed and developing countries responsible for promoting development and Heads of multilateral and bilateral development

Box 1.4	The MDGs

The Millennium Development Goals (MDGs) are the world's time-bound and quantified targets for addressing extreme poverty in its many dimensions-income poverty, hunger, disease, lack of adequate shelter, and exclusion-while promoting gender equality, education, and environmental sustainability. They are also basic human rights – the rights of each person on the planet to health, education, shelter, and security.

Goal 1 Eradicate Extreme Hunger and Poverty

Goal 2 Achieve Universal Primary Education

Goal 3 Promote Gender Equality and Empower Women

Goal 4 Reduce Child Mortality

Goal 5 Improve Maternal Health

Goal 6 Combat HIV/AIDS, Malaria and other diseases

Goal 7 Ensure Environmental Sustainability

Goal 8 Develop a Global Partnership for Development

institutions, meeting in Paris on 2 March 2005, resolve to take far-reaching and monitorable actions to reform the ways we deliver and manage aid' (OECD-DAC, 2005).

Notions of performance play a central role in the aid-effectiveness agenda. The Paris Declaration and its forerunners state the need for development agencies and national governments to harmonise their operational procedures. This has led to calls for bilateral and multilateral bodies to align their support with partner-country priorities and strategies through common arrangements for managing aid. This would have the benefit of reducing the burden on recipient countries of administration, information and relationship management.

The Paris Declaration also emphasises country ownership, without which national commitment to the MDGs and development processes would not be sustainable. Ownership can be established through policy tools such as the poverty-reduction strategy paper, or national development frameworks. Both of these involve aid-

receiving countries setting development aims and priorities, with donor efforts then being focused on directly supporting the realisation of these aims.

The third key element of the Paris Declaration is the notion of performance, articulated through the principles of Managing for Development Results (MfDR). Although results-based management has been established in some aid organisations since the 1990s, the movement has received increasing prominence through the aid-effectiveness agenda and the MDGs. The MfDR Sourcebook defines a 'result' as the output, outcome or impact – either intended or unintended, positive or negative – of a development intervention, and states:

> '... development stakeholders now recognise that the process of improving conditions in the world, a country, or an organisation is a process of change management. Defining clear results provides a better target for change. Periodically measuring results provides guideposts or markers that allow for correction to keep programs or projects on track toward their stated outcomes. Ultimately, better managing for results helps demonstrate more clearly whether development outcomes have been achieved' (World Bank, 2008).

A number of important factors are seen as contributing to better development results and performance, including:

- emphasis on continuous improvement at the institutional, programme and project levels

- devolution of management authority, ideally to national counterparts

- orientation towards service delivery and quality in terms of beneficiary needs and preferences

- participation of a wide range of stakeholders including ultimate beneficiaries in defining the ultimate results

- reform of finance and budget processes and systems to increase public transparency and accountability

- application of modern public management techniques.

There is clearly considerable overlap between these principles and those of the NPM processes in OECD countries, as listed above on page 32.

MfDR is now established as a global partnership providing countries and agencies with information on how to work together to manage for development results. It is directly reflected in the 2005 Paris Declaration. Regular performance assessment is an integral part of achieving the overall strategy and vision agreed for the sector. There are a number of principles which outline how such assessments should be undertaken. These principles are applicable at any level and within a variety of interventions (national, sector, program, project, and organization), and they influence the use of specific strategies and tools at various phases of national and development programming (World Bank, 2008). These include:

- **Principle 1** At all phases – from strategic planning through implementation to completion and beyond – focus dialogue on results for partner countries, development agencies, and other stakeholders.

- **Principle 2** Align actual programming, monitoring, and evaluation activities with the agreed expected results

- **Principle 3** Keep the results reporting system as simple, cost-effective, and user-friendly as possible.

- **Principle 4** Manage for, not by, results. Resources must be arranged to achieve outcomes.

- **Principle 5** Use results information for management learning and decision making, as well as for reporting and accountability

There are several other initiatives supporting the MfDR approach, including DFID's Multilateral Effectiveness Framework (MEFF), and Multilateral Organizations Performance Assessment Network (MOPAN). MEFF was established during 2003/04 to assess and monitor the organisational effectiveness of the international organisations that it funds (Scott, 2005). Given the scarcity of information on results from many programmes, and the difficulty of comparing them, MEFF uses results-based management theory to assess organisational effectiveness in terms of eight different systems including corporate strategy, resource management, quality assurance and results reporting.

Along similar lines, the Multilateral Organizations Performance Assessment Network (MOPAN) is a network of donor countries with a common interest in monitoring and assessing the performance of multilateral organisations. It carries out an annual survey focusing on multilateral partnership behaviour towards national stakeholders in developing countries and towards other international development agencies.

1.1.4   Performance approaches in NGOs and civil society

As outlined above, the NPM approach has been central to performance-oriented reforms of development aid agencies. Results-based management (RBM) has been the prominent approach among donors and multilaterals. Although such approaches have been 'pushed' on to implementing non-governmental organisations, it is interesting to note the different principles and approaches that have seen take-up among NGOs and civil society organisations.

'The number and size of NGOs has reached levels never seen before. This is mainly a result of the decentralisation of the state and deregulation of its services that began in the 1980s and saw a shift in the role of NGOs and businesses alike. This saw the number of NGOs registered in the 30 member countries of the OECD almost double from 1,600 in 1980 to 2,970 in 1993, with a concomitant doubling of spending.'[4]

The growth in civil society has been accompanied, as in the public sector, by an increasing focus on accountability and performance. Traditionally, ideas about the relative performance of NGOs rested on perceptions of their ability to meet social needs, in contrast to the suggested failings of the market and the state. NGOs were seen as a 'voluntary' or 'third' sector that was less bureaucratic and hierarchical than government and more altruistic than the 'free' market focused on financial profit (Korten, 1990; Brett, 1993). Underlying this was the belief that NGOs are seen to uphold high levels of altruistic motivation (Brett, 1993), and so would work for less than the going rate of remuneration and would engage in activities in which no firm would invest.

NGOs were also seen as more flexible than large and hierarchically structured government bureaucracies and firms (Edwards and Hulme, 1995), with decisions and activities based on close interpersonal relations, and unconstrained by bureaucratic codes of behaviour. Because of this, NGOs were seen as able to adapt

their project priorities to suit local conditions. Public interest in and contributions to the work of NGOs has increased concurrently with governments seeing them as a convenient and inexpensive means of delivering public services (ibid), leading to the growth in both the number and size of NGOs.

Work by Accountability and Keystone, two leading organisations in the field of transparency and accountability, highlight two different waves of NGO performance scrutiny (Accountability and Keystone, 2007). The first wave of was a response to the growing role of NGOs in social and economic development as a result of the decentralisation of state services during the 1980s. The involvement of NGOs in the delivery of public goods which was previously the responsibility of the state led to a focus on the effectiveness of NGO performance. The growth of NGOs into multinational organisations with significant operating budgets challenged some of the traditional perspectives on their relative performance advantage.

Citizen organisations also began to examine their own practices and a range of self-regulatory codes of conduct, quality standards and accountability mechanisms were developed. The tools and standards took on a technical focus in trying to improve the effectiveness of service delivery (Ebrahim, 2003), and a number of NGOs have developed internal quality assurance and performance mechanisms. Legal frameworks, such as the Charities Commission in the UK, were also established around compliance codes focusing on specific elements of performance.

The second wave of scrutiny dates back to the late 1990s, and was a response to the advocacy and lobbying activities of NGOs. As NGOs took on a greater role in global governance, there was an increasing tension between the service-oriented role as envisaged by states, and the political role whereby NGOs become necessary voices to 'speak truth to power' and to point out deficits within the state and private sectors. This has led to numerous questions about the legitimacy of citizen organisations in representing the poor and disadvantaged, and their 'right' to attack governments and businesses.

International development NGOs

As in the wider third sector, NGOs working in international development and relief have in the last few years been paying increasing attention to demonstrating their performance and the impact of their work on those they claim to help. Approaches to managing NGO performance can be grouped into four categories (BOND, 2006).

1 Statutory regulations are legal requirements of NGOs in the country and region in which they operate. They include laws and conventions enacted by government and multilateral bodies such as the UN. The overriding regulation guiding NGOs is the range of human rights law and international humanitarian law, on which other approaches are based.

2 Voluntary principles and codes are performance standards that are essentially self-regulation. Examples include the Red Cross Code of Conduct and the Humanitarian Accountability Partnership International's current initiative to develop a standard of accountability and quality management that focuses on intended beneficiaries.

3 Organisational management and measurement tools assist NGOs in implementing and adhering to statutory regulation and normative principles as well as being used for general organisational development. These include individual organisational approaches, such as ActionAid's Accountability, Learning, and Planning System (ALPS); proprietary approaches, such as Investors in People; or open-access models, including Projet Qualité and Synergie.

4 Evaluation and verification processes can be included in the other categories above, and involve an added assurance mechanism in the form of external evaluation, or financial and social audit, which may or may not lead to certification or accreditation. For example, NGOs implementing the People In Aid Code are independently audited (based on the AA1000 Assurance Standard) against the principles of the Code. For more on the code, see later section on organisational approaches to performance in the humanitarian sector.

Despite this range of mechanisms, performance transparency remains a major challenge for NGOs:

'Transparency about programmes and operations, as opposed to basic financial accounts, is not easy when there is pressure on citizen organisations to demonstrate results in short timeframes, and in quantitative form. Citizen organisations are locked in a dilemma of knowing that a lack of transparency is both damaging the sector, and contradictory to its values. At the same time they fear that a greater openness about performance by the

sector will lead to a negative reaction by the media, donors, and the general public' (Keystone and Accountability, 2007, p. 4).

1.1.5 Cross-organisational and sector-wide performance in the public and social sectors

Sector-wide performance approaches have been used in some private-sector operations – for example, by developing financial indices relating to the performance of all the firms working in a particular sector such as banking or mining. They might also be used to look at particular aspects of performance – for example, corporate social responsibility across energy companies, or safety in the nuclear industry.

However, it is in the public sector that such approaches have seen greatest degree of take-up. Arguments for system-wide approaches have been presented as fundamental to understanding public- and third-sector performance in a range of settings.

> 'A single successful agency does not make a successful public policy system; a well performing hospital does not amount to an effective health system; a well-performing school does not mean that a well-performing education system is in place... for that a vast network of organisations must work in synergy with each other to achieve the desired public policy outcomes and create net public value' (Bouckaert and Halligan, 2008).

The prominence of inter-organisational networks has been described as a defining characteristic of twenty-first-century managerial and organisational life, whether in government, business or civil society (Bourgon, 2008). System-wide approaches help to recognise the reality of multiple organisational relations within and beyond any single set of actors or sectors, and the importance of working through networks, partnerships and other coordinating mechanisms. Where they are successful, such approaches follow the chain of activities among actors leading to overall goals and outcomes. System-wide approaches are challenging for both political and methodological reasons, but it has been argued that they are the best way to enable change at an operational and political level (Bourgon, 2008).

In the global public sector and civil society, there are increasing numbers of approaches and mechanisms that appear to take a system-wide approach. Some of these take particular issues – such as accountability – and review the performance of a range of organisations. For example, One World Trust's Global Accountability Report looks at different kinds of international organisations; Transparency International's Corruption Index looks at corruption at the level of nation states. Also working at national level are tools such as the Human Development Index, used by UNDP since 1990, and the Commitment to Development Index of the Centre for Global Development, which looks at donor agencies' efforts to enable development.

The global civil society network, CIVICUS, looks at countries according to a civil society index. There are also increasing numbers of reports which look at the 'state of' different issues. Many of these are commissioned and implemented by UN agencies, for example, UNEP on forests, UNHCR on refugees, UNICEF on mother-and-child poverty, UNISDR on disaster-risk reduction. Others are implemented by civil society organisations – for example, Amnesty International's State of Human Rights. Global peace operations have been reviewed on behalf of the Department of Peacekeeping Operations (DPKO) on a biennial basis, using a number of key performance indicators across operations.

Most of these approaches use an implicit or explicit balanced scorecard approach, based on identifying a number of different aspects of performance in relation to an overall goal, and comparing the performance of organisations or countries against them. Such benchmarking is usually done relative to organisations' own historical performance or to other comparable organisations. The frameworks used tend to be composites of approaches, drawing on different aspects of different models.

The influence of such approaches in improving performance is not well documented – for example, the same countries have topped the Transparency International Corruption Index for some years now, and likewise the UNDP Human Development Index. However, the value of comparative approaches in system-wide assessments has been questioned by a number of analysts. The assumption is that gathering and reporting comparative information creates a powerful incentive for improving and learning from others. However, this also assumes that what appears to work in one organisation can be readily and easily transferred to another. A study of performance benchmarking among not-for-profits finds that some important ideas *can* be imported into organisations through such means, but more frequently they are picked up through cross-organisational networking and information exchange.

While most comparative approaches do not assess performance across a sector, one exception is the work of the World Health Organisation (WHO) in developing a performance assessment framework for health systems at country level (Murray and Frenk, 2006). This framework recognises sector-wide performance as an interactive process in which social outcomes, such as health, responsiveness and fairness, are the most valued objectives. Recognising the role of context and available resources, the WHO framework suggests that performance should be assessed relative to the worst or best that could be achieved given a particular set of circumstances (ibid).

Performance indicators typically used in both the public and third sectors are listed in Box 1.5.

Box 1.5	Indicators of performance used in public and not-for-profit sectors

Performance management in the public and NGO sectors is commonly framed in terms of the 'three Es' of economy, efficiency, and effectiveness.

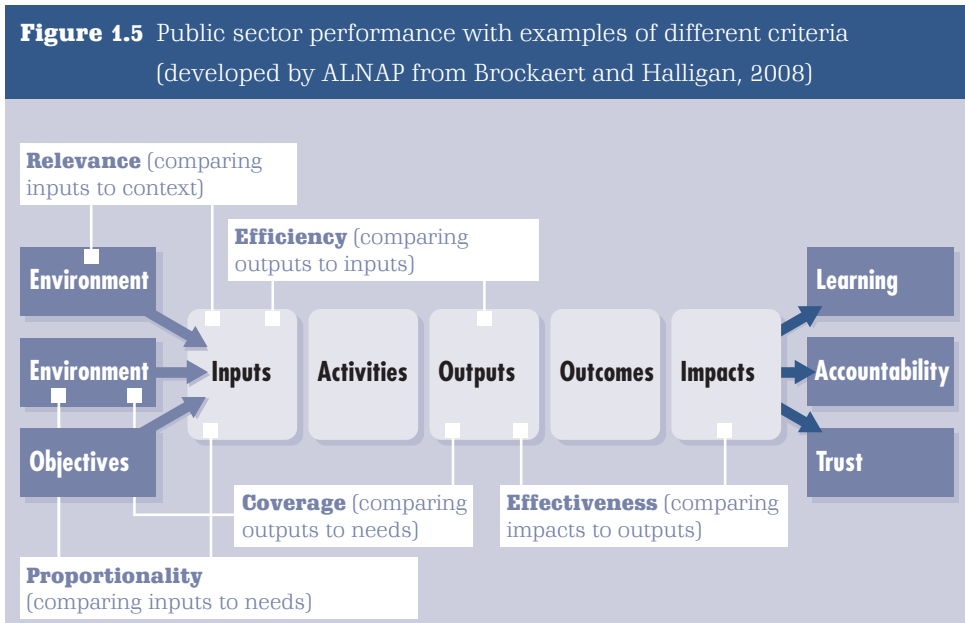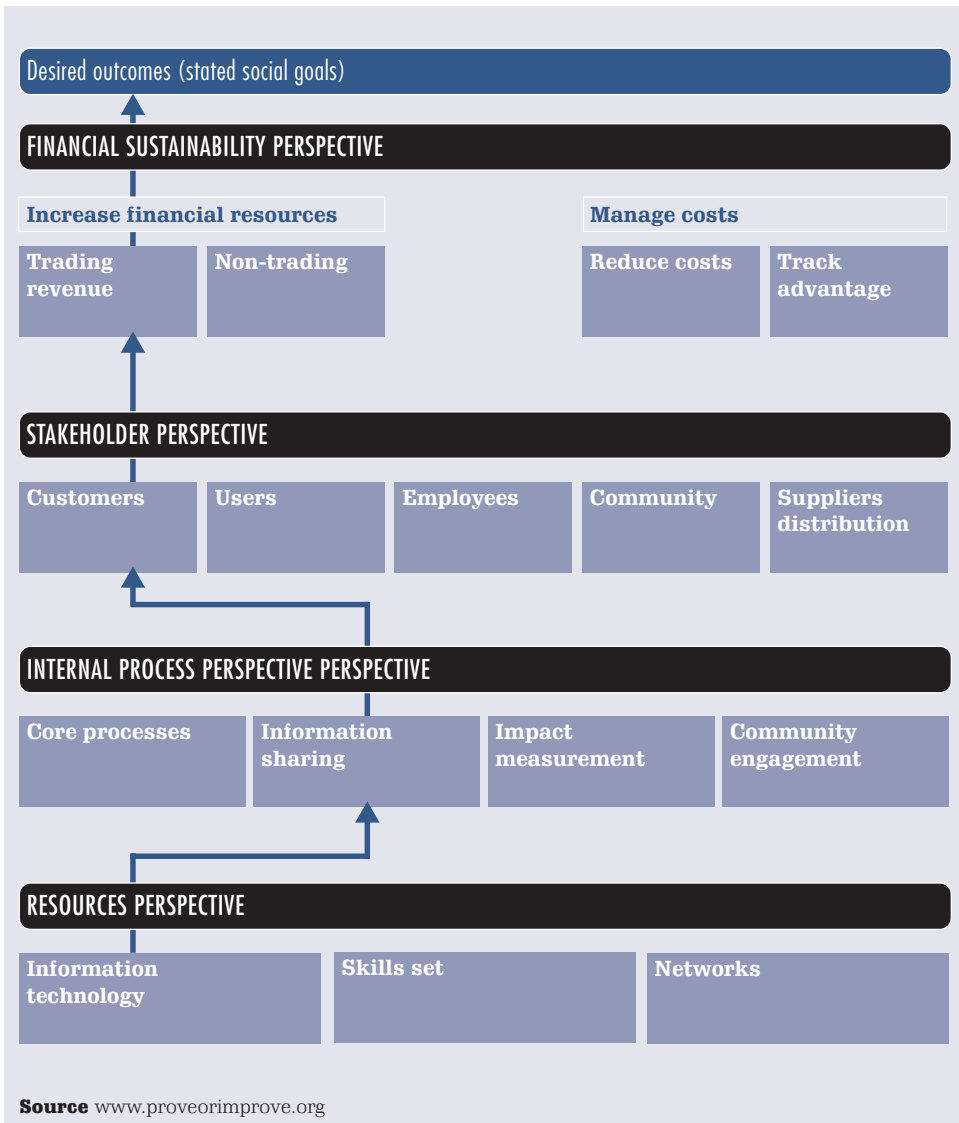| Economy of operations	Efficiency of management	Effectiveness of strategy
• Procurement and delivery of inputs • Human, physical and financial resources • Quantity and quality • Cost element • Timeliness	• 'Optimal transformation (activities) of inputs into outputs' (UNDP) • Utilisation of means to achieve results and objectives • Rational use of resources • Least costs for maximum results/return • Activities with respect to results • Work planning and timelines	• The extent to which a programme or project achieves its immediate objectives or produces its desired outcome • Achievements of results, objectives or goals • Focus on changes in lives of target groups, beneficiaries or clients • Medium- and long-term perspective

Figure 1.5 presents an illustrative model of how these different elements might fit together in a performance framework. This illustrates the full span of performance from instigating factors such as environmental triggers, through to the final goals of learning, accountability and building trust. Distinct aspects of performance, including proportionality, coverage, relevance, efficiency and effectiveness can be determined by comparing different sorts of data at different points in the performance span or results chain. For example, efficiency can be assessed by comparing output data with data from inputs, and so on.

Figure 1.5 Public sector performance with examples of different criteria (developed by ALNAP from Brockaert and Halligan, 2008)

The trend in the private sector towards an increased focus on stakeholder consultation and internal capacities and resources is reflected in the new wave of 'balanced' models. However, this is generally not *explicitly* addressed in public-sector indicators, or in results frameworks such as the one presented above.

A more balanced approach to public-sector and NGO performance could help in combining diverse approaches. For example, the balanced scorecard developed for social enterprises brings together outcomes, financial issues, stakeholder perspectives, internal processes and resources/capacities (Figure 1.6). This also has a 'performance logic' shown by the dashed arrows, which illustrate how – in this case – better IT can contribute to high level outcomes.

Figure 1.6 Balanced scorecard for social enterprises

Desired outcomes (stated social goals)

FINANCIAL SUSTAINABILITY PERSPECTIVE

Increase financial resources

| Trading revenue | Non-trading |

Manage costs

| Reduce costs | Track advantage |

STAKEHOLDER PERSPECTIVE

| Customers | Users | Employees | Community | Suppliers distribution |

INTERNAL PROCESS PERSPECTIVE PERSPECTIVE

| Core processes | Information sharing | Impact measurement | Community engagement |

RESOURCES PERSPECTIVE

| Information technology | Skills set | Networks |

Source www.proveorimprove.org

1.1.6 Overview: How well is performance working?

This section offers a framework for the ideas presented so far in this chapter. Borrowing from reviews of performance in the development sector and further afield (Bouckaert and Halligan, 2008; Obser, 2007), there are three elements of performance systems:

1 developing performance frameworks, collecting performance data and processing them into information – collectively 'measuring'

2 integrating information into documents, procedures and stakeholder discourses

3 using performance management in a strategy of improving decision-making, results and accountability.

In terms of these elements – measuring, integrating and using – and with the addition of 'limitations', the two authors identify four distinct models or stages of performance (Table 1.1).

Table 1.1 Four models of performance

	MODEL 1	MODEL 2	MODEL 2	MODEL 4
	Traditional/ pre-performance	Performance administration	Management of performances	Performance management
MEASURING	Intuitive	Administrative data registration	Specialised performance measurement systems	Hierarchical performance measurement system
INTEGRATING	None	Some	With different systems for specific management functions	Systematically integrated
USING	None	Limited	Disconnected	Coherent, integrated, comprehensive, consistent
LIMITATIONS	Function unawareness	Ad hoc selective	Incoherence	Sustaining complex system

Source Brouckaert and Halligan, 2006

1   Traditional/pre-performance models relate to those organisations and sectors in which there are no systems for gathering data or sharing it, and performance systems are intuitive and informal. Data will be weak, not integrated into wider organisational communications, and unusable.

2   Performance administration models reflect a commitment to and expectation of performance assessment and management, but the system may not be explicit or well developed and application is often ad hoc. A generalised call for performance translates into another administrative procedure that may be part of an administrative and legal setting, rather than a managerial or policy context.

3   Management of performances models apply where management and performance have been linked but the connection between them is underdeveloped and a number of different systems operate for different, unconnected management functions. This results in a diverse range of managements with distinct performances, which can lack consistency and coherence, even though some functions might be highly developed driving improvements and reforms in other functions.

4   Performance management models describe a system that is coherent, integrated, consistent and comprehensive. This involves a solid performance system which is used in a coherent strategy of management improvement. Different performance assessment systems are used for different purposes, with coordination across the different systems. This model requires an explicit attention and focus on policies and practices for assessment as well as managing different functions and their performances.

For all sectors, as discussed above in Section 1.1, reviews of performance systems identify widespread concern, dissatisfaction and cynicism around approaches to performance. In the private sector it has been suggested that significant numbers of organisations decide not to implement a performance system because they see no advantage (Neely et al., 2007), or that the resources required for implementation were simply insufficient, leading to weak and poorly utilised systems.

Problems in the public and third sectors appear even greater. Performance management here faces a number of distinct problems relevant to humanitarian operations. In the public sector, performance management has been found to produce a range of 'perverse incentives' and 'gaming responses' by managers.

Perhaps the most widely cited example is the attempts of the UK National Health Service to meet government waiting-list targets by distorting clinical priorities. According to one report (NAO, 2004), more than 50 per cent of consultants reported treating non-urgent cases before serious ones in order to meet waiting-list targets.

Despite the understanding that no system can be all things to all people, public-sector performance mechanisms have been used to serve multiple purposes – for example, both control and learning, innovation and improvement. The various uses of performance information may not be completely compatible with one another, or may require different types or levels of result data to satisfy their different needs and interests. Balancing these different needs and uses without over-burdening the performance-management system remains a major challenge.

Focusing on efforts in development agencies highlights fundamental problems about the very nature of performance systems and their goals. A review of results-based management (RBM) among donor agencies (Meier, 2003) found a frequent emphasis on activities and outputs, despite the fundamental problem this raises with the rationale for RBM. A more recent study of RBM among UN agencies (OIOS, 2008) identifies concerns at conceptual and practical levels, as follows.

• The formalistic approach to codifying how to achieve outcomes, inherent to RBM, can in fact stifle the innovation and flexibility required to achieve those outcomes.

• The original design of RBM is inadequate, statements of results are often vague, and the determination of success does not lend itself to impartial, transparent and precise measurement.

• Accountability is not based on a review of outcomes but instead on ascertaining that there is no negligence, misconduct or breach of rules and regulations.

• Results-based management has made no contribution to strategic decisions.

• Outcomes are influenced by multiple actors and external risk factors outside UN control.

• Results-based management is a cultural rather than technical issue.

- The introduction of results-based management has been treated as an addition to the rules and procedural requirements that govern inputs, activities, monitoring and reporting, with no relaxation of the existing volume of regulatory frameworks relating to financial, programmatic and human-resource management.

- The rules and regulations for programme planning, budgeting, monitoring and evaluation have blurred the distinction between the separate roles of evaluation versus monitoring and of independent evaluation versus self-evaluation.

- Performance measures frequently lack baselines and targets, and many are not regularly tracked.

- Many of the results planned for have been expressed in a self-serving manner, lack credible methods for verification and involve reporting that rests upon subjective judgement.

- 'Expected accomplishments' largely relate to individual sections or divisions, with no target-setting or measurement for objectives that transcend divisions or departments or seek to capture longer-term objectives of the UN as a whole.

- Although aspirational results are used to justify approval of budgets, the actual attainment or non-attainment of results is of no discernible consequence to subsequent resource allocation or other decision-making.

- The metrics do not exist for systematic determination of the efficiency and effectiveness of the organisation.

On this basis, it seems that performance management in international development has not moved significantly beyond the 'performance administration' model, and may in fact be stuck there because of the pre-existing and unchanged accountability relationships. The lack of harmonisation of performance approaches has meant that any real focus on performance has been drowned out by information demands. Much of this information is not used in any practical sense, but seems to be collected primarily for accountability purposes.

Despite these issues, it is acknowledged that the performance agenda is here to stay, and that more work is needed to improve and streamline it, and make it part of the

culture of international organisations (OIOS, 2008). Somewhat ironically, perhaps, the results-based agenda does not appear to need results to fuel or sustain commitment to it (Bouckhert and Halligan, 2008).

1.2   Performance in international humanitarian action

How does all of this relate to international humanitarian work? Discussions on performance management within ALNAP date back to the 10th Network Meeting in 2001, which highlighted the potential value of understanding private-sector frameworks for performance and quality management. The rationale at that time was that attempts to improve humanitarian performance had to date focused on improvements in the context of projects and programmes, such as by: establishing monitoring and evaluation systems; using project-based contractual obligations, codes and standards; and exploring mechanisms for ensuring accountability to stakeholders.
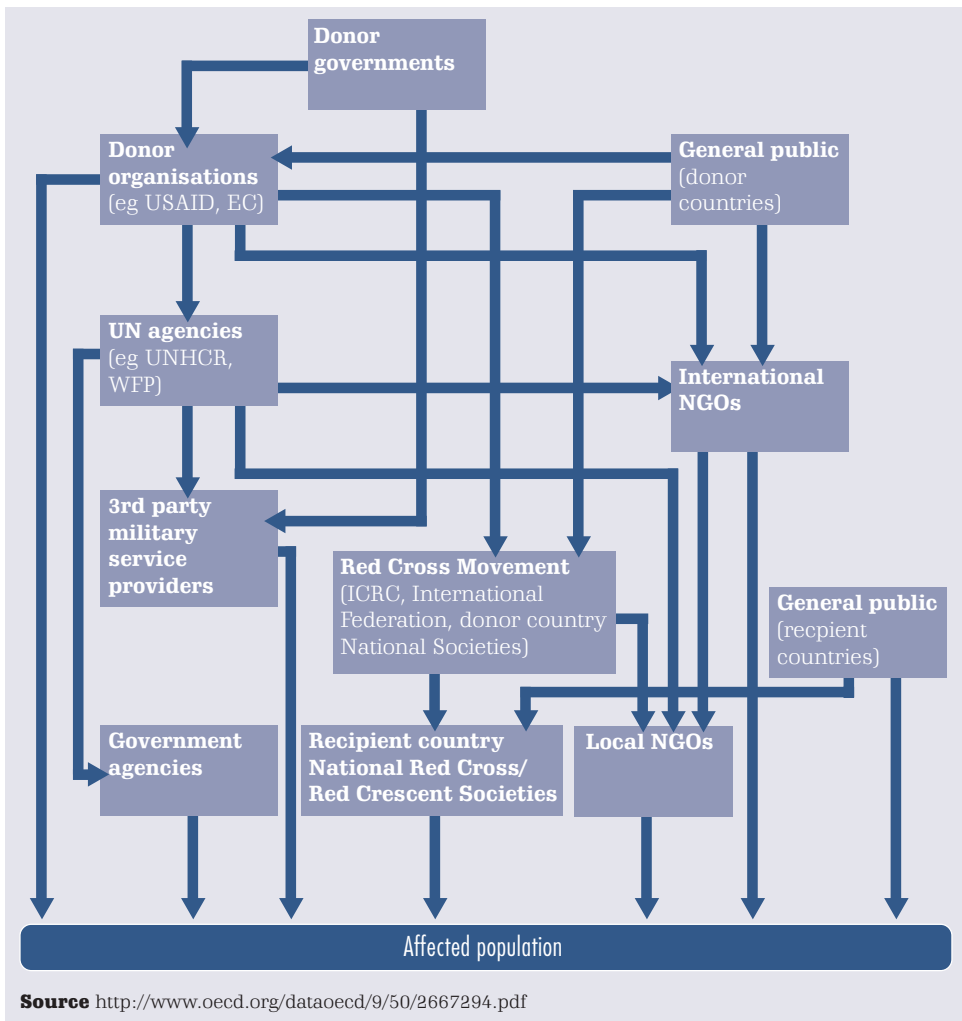
However, these discussions and reflections also highlighted the lack of significant institutional or cross-institutional use of more rounded performance models such as EFQM or Balanced Scorecard. A review by Andre Griekspoor (Griekspoor, 2001) identified gaps in quality and accountability tools arguing that holistic frameworks such as EFQM could help to broaden perspectives on performance – from project-level to organisational and system-wide – and that such approaches could also be used to embed humanitarian improvement initiatives in a broader framework.

The concept of performance has remained a guiding light in much of ALNAP's work, as it has for all of the other 'quality and accountability' initiatives within the sector. Before considering performance itself – what it consists of and how is it assessed and managed across the sector – Section 1.2.1 looks first at the humanitarian system itself.

1.2.1 The nature of the humanitarian system

As shown in Figure 1.7, the humanitarian system has been described as a complex web of interdependent relationships that cut across institutional and geographical boundaries (Hallam, 1998; Stoddard, 2007), and the features of the system arise from the nature of these interactions (Graves and Wheeler, 2006). The humanitarian sector consists of multiple agencies and departments with many different goals and objectives.

Figure 1.7 The complexity of aid flows in crises

Source http://www.oecd.org/dataoecd/9/50/2667294.pdf

There are five core sectors within humanitarian response: health services; water, sanitation and hygiene promotion; food security and nutrition; aid items (food and non-food), and shelter and settlement. Many humanitarian operations also feature a wide range of other activities, some of them characterised within broad themes cutting across the core sectors. These themes include: protection, gender, education, agriculture, psycho-social/mental health support, capacity strengthening, income generation, infrastructure rehabilitation/reconstruction, human rights approaches and support to the re-establishment of the rule of law.

There are three other key international sectors with which relief work can overlap:

1 disaster-preparedness work, especially in countries with a known risk of disasters, and increasingly given issues of climate change – investment in disaster-risk management and reducing the vulnerability of 'at-risk' communities is often more cost-effective than intervening after disaster has struck. Increasingly, this means engaging with development and environmental agencies

2 peacekeeping missions, which work alongside and sometimes closely with humanitarian agencies of the UN and international and national NGOs, in countries with ongoing peacekeeping operations

3 human rights efforts, in which national and international agencies are involved in monitoring protection issues and abuses in settings where relief agencies are delivering aid. For example, the rights monitoring undertaken by Amnesty International. This is distinct from relief agencies own rights-based work.

The range of activities undertaken within humanitarian responses is now extremely broad. This reflects the heterogeneity of the system and the actors within it. However, the following humanitarian principles form a shared framework guiding humanitarian activity.

• Humanity – that humankind shall be treated humanely in all circumstances by saving lives and alleviating suffering, while ensuring respect for the individual.

• Impartiality – the provision of humanitarian assistance according to need alone, and not based on nationality, race, religion, or political point of view.

- Independence – that humanitarian agencies must formulate and implement their own policies independently of government policies or actions.

The Code of Conduct for the International Red Cross and Red Crescent Movement and NGOs in Disaster Relief (often referred to as the Red Cross Code of Conduct) was drawn up in 1994 by seven of the largest humanitarian agencies. It uses humanitarian principles as the basis for specifying how humanitarian work should be undertaken (Box 1.6). Over 300 organisations and agencies have agreed to abide by the Code.

More recently, humanitarian principles formed part of the preamble to the 'Principles and Good Practice of Humanitarian Donorship', established by all 23 members of the OECD's Development Assistance Committee. This initiative defined the objectives of humanitarian action as:

Box 1.6	Principal commitments of the Red Cross Code of Conduct

- The humanitarian imperative comes first.

- Aid is given regardless of the race, creed or nationality of the recipients, and without adverse distinction of any kind. Aid priorities are calculated on the basis of need alone.

- Aid will not be used to further a particular political or religious standpoint.

- We shall endeavour not to act as instruments of government foreign policy.

- We shall respect culture and custom.

- We shall attempt to build disaster response on local capacities.

- Ways shall be found to involve programme beneficiaries in the management of relief aid.

- Relief aid must strive to reduce future vulnerabilities to disaster as well as meeting basic needs.

- We hold ourselves accountable to both those we seek to assist and those from whom we accept resources.

- In our information, publicity and advertising activities, we shall recognise disaster victims as dignified human beings, not hopeless objects.

'to save lives, alleviate suffering and maintain human dignity during and in the aftermath of man-made crises and natural disasters, as well as to prevent and strengthen preparedness for the occurrence of such situations' (GHD, 2003).

Some humanitarian agencies would take issue with elements of this definition, in particular those agencies for whom working to prevent or strengthen preparedness would work against the principle of independence. As the OECD-DAC acknowledges:

'... in situations where the state is a party to armed conflict or has otherwise abrogated responsibilities for the safety and welfare of civilians on its territory, full association... with respect to partner government ownership... would compromise core principles of neutrality, impartiality and independence of humanitarian action' (OECD-DAC, 2008).

1.2.2   Defining and introducing humanitarian performance

A recent Oxfam report on fundamental changes needed in humanitarian assistance in response to climate change usefully sums up how aid should be delivered: 'People affected by emergencies deserve aid that... is relevant, good quality, and well-managed... is accountable, with mechanisms to challenge failure and abuse; builds durable solutions; and is sufficiently resourced' (Oxfam, 2009).

Humanitarian performance could be therefore be seen as 'performance of those activities and efforts that seek to fulfil the principles and objectives of humanitarian action'.

Combining these perspectives, a useful working definition of humanitarian performance might be as follows:

"...the effective collective performance of a *complex system of international, national and locally-based organisations*, which works to *save lives, alleviate suffering and maintain human dignity* both during and in the aftermath of man-made crises and natural disasters, as well working to prevent and strengthen preparedness for the occurrence of such situations. Effective

> performance means undertaking work *in ways that are consistent with humanitarian principles*, mobilising and deploying sufficient financial, material and human resources *in ways that are relevant, well-managed, accountable, impartial, durable and ensure good quality…"*

Contained in this definition are the different elements of humanitarian performance now seen as important for the sector. These have come to the fore since the 1990s, in response to a number of now-familiar criticisms and concerns. The overall implication of these critiques has been that relief agencies are largely self-serving, and so form part of the problem rather than the solution (Leader, 2000).

In response, a search for new ways of working has generated a variety of conceptual and operational efforts and initiatives to:

- establish accountability through standards and codes

- improve performance through information and management techniques such as needs assessment, project-cycle management, and evaluations

- emphasise local capacities, linking relief to development, and disaster-risk reduction

- focus on the rights of affected people and on protection.

Perhaps unsurprisingly, just as the system is fragmented across multiple levels and actors, so too are the mechanisms and initiatives aiming to monitor and improve performance within it.

A number of broad levels can be articulated, which demonstrate the potential depth of performance approaches in the sector. These are illustrated in Box 1.7 overleaf.

To contribute to understanding performance approaches and challenges in more detail, the rest of this section explores the range of initiatives in the humanitarian system by relating them specifically to the essential workings of different aspects of the system. First, the context of performance needs is considered in terms of global and local situation analysis, needs assessment and resources mobilised across the system (Section 1.2.3), followed by a review of approaches to understanding project and programme performance (Section 1.2.4). Next is an analysis of approaches to

---

Box 1.7   Potential depth of humanitarian performance assessment

- Specific project or programme level (e.g. An NGO shelter programme in the tsunami)

- Organisational within operations (e.g. all of an NGO's efforts in Mozambique) or across operations (e.g. UN agency efforts across the tsunami) or globally (e.g. donor efforts in humanitarian response over time around the world)

- Sectoral within an operation (e.g. emergency health work across all relevant agencies post-Pakistan earthquake)

- Multi-sectoral within an operation (e.g. health and nutrition in DRC) or across operations (e.g. food programmes in the Horn of Africa)

- System-wide within a particular operation (e.g. the entire relief system in Darfur) or across operations (e.g. overall response to the Great Lakes crisis)

- Sector at a global level (e.g. emergency water and sanitation globally)

- Key theme at an operational level (rights in Darfur), across operations (e.g. LRRD in the tsunami), or globally (e.g. donor performance, gender, rights or innovations across the system)

- System-wide, globally

- Beyond the system: how actors outside the system influence or are influenced by the system

---

performance at the operation-wide level, including health and nutrition tracking systems, joint evaluations and beneficiary assessments (Section 1.2.5). Organisational performance mechanisms (Section 1.2.6) and approaches to system-wide performance (Section 1.2.7) are then reviewed. A visual overview of these different aspects is provided opposite.

1.2.3   Situation analysis: contextualising humanitarian performance

To understand performance, its context must also be understood in terms of needs and the relative availability of resources. The context of humanitarian performance has four broad dimensions, as considered here: the global context of relief, operational contexts, needs assessments and funding.
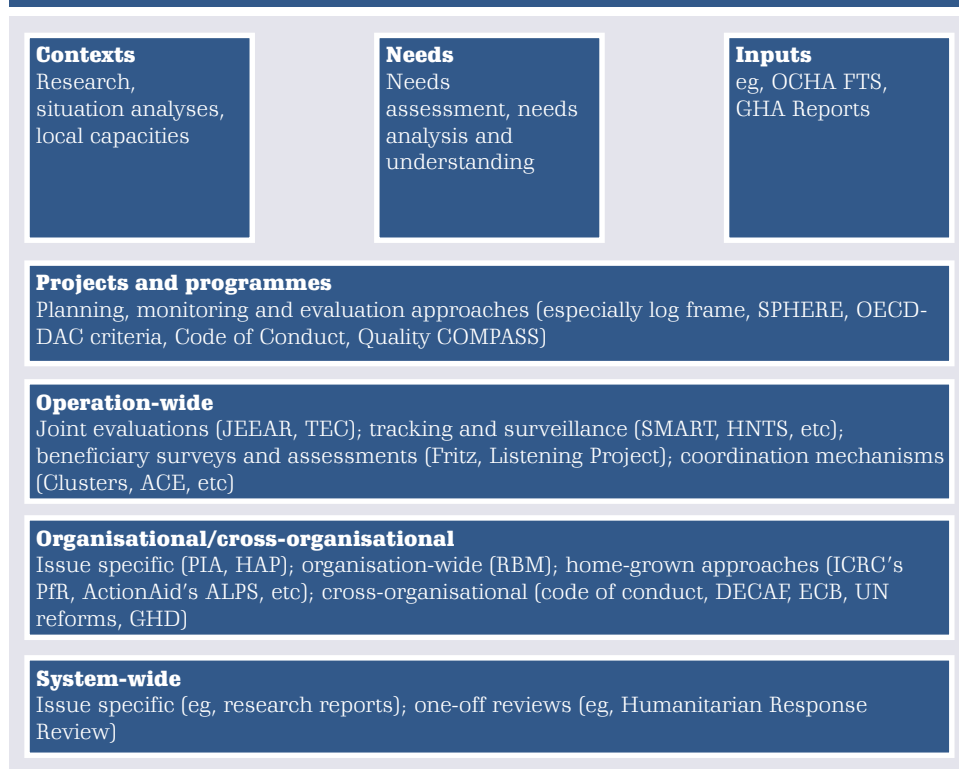
Global contexts

As highlighted by previous ALNAP work, expectations of international humanitarian aid have increased dramatically in the last twenty years, with humanitarian relief now featuring more prominently in world politics. Aid is now increasingly regarded as a right and an obligation under international law – and as an element of a global welfare system focused on those affected by war and disaster. Realising this ideal would depend on changes in global politics as well as on humanitarian practice itself.

However, global welfare may not be a realistic expectation, and there may be greater scope for change within the 'inner world' of the global aid system (in terms of learning, innovation and capacities) than in the outer-realm issues of politics.

Political influence can be seen as a cause of distortion and inefficiency in humanitarian work, especially given the differential allocation of resources according to criteria

Figure 1.8 An overview of the performance mechanisms reviewed

Contexts
Research, situation analyses, local capacities

Needs
Needs assessment, needs analysis and understanding

Inputs
eg, OCHA FTS, GHA Reports

Projects and programmes
Planning, monitoring and evaluation approaches (especially log frame, SPHERE, OECD-DAC criteria, Code of Conduct, Quality COMPASS)

Operation-wide
Joint evaluations (JEEAR, TEC); tracking and surveillance (SMART, HNTS, etc); beneficiary surveys and assessments (Fritz, Listening Project); coordination mechanisms (Clusters, ACE, etc)

Organisational/cross-organisational
Issue specific (PIA, HAP); organisation-wide (RBM); home-grown approaches (ICRC's PfR, ActionAid's ALPS, etc); cross-organisational (code of conduct, DECAF, ECB, UN reforms, GHD)

System-wide
Issue specific (eg, research reports); one-off reviews (eg, Humanitarian Response Review)

other than need. More difficult to analyse and track is the degree of politicisation of relief by Western states, the role of earmarking or tying aid to particular forms (especially food aid) or sources (usually organisations from the nation state in question). Here, an understanding of context requires attention to: foreign policy and domestic political interests of donors; historical, cultural and geographic ties that shape resource flows; and agency-specific concerns with marketing and profile. Media coverage of emergencies is important here, in terms of the balance of coverage given to different crises.

Operational contexts

Understanding performance also requires an understanding of operationally specific context. This can be viewed in terms of the four key elements of any disaster or crisis (Cosgrave and Herson, 2008, p 182):

• disaster type;

• the pre-existing nature and circumstances of the affected population;

• local, national, and international response – including the local and national capacity to respond; and

• the areas affected.

The context in which humanitarian operations take place often has a very significant influence on their performance. An operation in which agencies have unimpeded access to all of the affected people and areas will be able to achieve more than an operation in a different context, for example.

In general, a wider historical and geographical perspective can enable better understanding of such issues as why the relative power and vulnerability of different groups changes over time, and how the fortunes and activities of one group can affect others. This can help agencies to monitor and assess needs for assistance and protection more effectively (HPG, 2003)

While any assessment of performance should take full account of the context of an operation, such understanding does not routinely or effectively inform humanitarian operations. Often it is left to subsequent evaluations to point out the mismatch

between aid programmes and the context in which they are delivered. Increased engagement of local and national actors could bridge this gap, but are currently under-utilised in such efforts.

Needs assessments

As noted in a previous ALNAP review, needs assessments are a 'fundamental flaw of the humanitarian system' (Sandison, 2006), and moreover there is a clear difference between 'assessing needs' and 'analysing and understanding needs'.

The tendency remains for agencies to 'assess' needs in terms of what they can deliver, rather than in terms of what affected people most want. However, especially since publication of Darcy and Hoffman's influential study (Darcy and Hoffman, 2003), considerable effort has been made in this area by UN agencies, donors and researchers.

Substantial advances have been made in the process of developing a common framework for assessing and making comparisons of basic needs across operations and socio-economic contexts. The potential of needs assessments, done effectively, is that they can ensure efforts are more relevant and appropriate, and therefore set an initial evidence-based 'marker' for assessing performance.

There are several initiatives combining needs assessments and situational analysis in different ways, such as: WFP's Strengthening Emergency Needs Assessment Capacity Project[2]; the IASC's work on strengthening analysis and presentation of humanitarian needs in the Consolidated Appeal Process (IASC, 2007); and ECHO's Global Needs Assessment for assessing relative needs at a macro-level for approximately 135 countries, and which includes an index of 'Forgotten Crisis Assessment' (ECHO, 2004). There have been a number of attempts to develop joint needs assessments, but none to date have proved particularly successful.

Monitoring and reporting of funding flows and other inputs

The final part of the context of humanitarian performance is the resources mobilised to support and fund activities. There are difficulties here with fragmented and incomplete financial data, so it can be hard to know how much has been spent on humanitarian assistance, and how much has actually arrived on the ground. The two

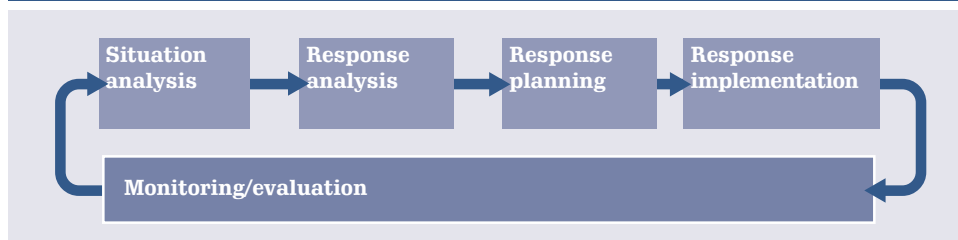principal sources of financial data are the OECD-DAC databases and the UN OCHA Financial Tracking System (FTS).

The DAC databases cover official development assistance (ODA), other official flows (OOF) and private funding (foreign direct investment, bank and non-bank flows) from DAC and other donors as well as multilateral organisations. OCHA's FTS captures voluntary statements of contributions to humanitarian assistance from over 60 countries, validated by the agencies receiving the contributions (Development Initiatives, 2008).

Data from these two sources are analysed, approximately every 18 months, by the Global Humanitarian Assistance programme of Development Initiatives. This innovative research programme aims to improve the efficiency, effectiveness and coherence of humanitarian response by further increasing access to reliable and transparent data on humanitarian assistance. Key challenges in this work include the lack of information on contributions from some sources, and particularly on domestic responses to crises within affected countries. Consequently, decision-making on international resource allocation is reduced in effectiveness and efficiency.

However, the key point here is the very considerable variation between operations in the level of resources provided in relation to the level of need (for examples, see Telford and Cosgrave, 2006), which inevitably has a major influence on performance.

1.2.4 Performance at project and programme level

Within the humanitarian system, performance is still mainly considered in the context of projects and programmes, as the primary contractual and organisational units of aid delivery. Within these, the key means for establishing and assessing performance remain effective planning, monitoring and evaluation techniques, and programmes increasingly use approaches based on the 'project cycle' (Figure 1.9). In reality, the first three components are conflated, with situation analysis running parallel to response analysis, planning and fundraising. However, for analytical purposes, it is worth considering them separately. Situation analysis is considered in the previous section (1.2.3). Therefore, this section concentrates on response analysis, planning, monitoring and evaluation, and their different roles in enabling

Figure 1.9 Idealised process of aid delivery in crisis contexts

Situation analysis	Response analysis	Response planning	Response implementation

Monitoring/evaluation

performance to be assessed and improved. The section ends with a brief overview of tools and frameworks to help with project and programme performance.

The role of planning in improving performance

Planning efforts in response to crisis involves first analysing the situation (as discussed in Section 1.2.3), and then analysing and planning the response. This usually means establishing common objectives and strategies, to ensure that operational agencies are working towards clear goals, and in line with agreed principles. The logic of setting objectives, and by extension assessing performance against them, is implicit in most programmes and projects. This demonstrates an implicit use of 'management by objectives', although the objectives are usually set at the level of activities (such as delivering a given amount of food) or outputs (such as feeding a given number of people). The span of performance is therefore limited by such approaches, to judging efficiency rather than effectiveness.

The primary tool used for this purpose is the much maligned logical framework analysis (LFA) (Bakewell and Garbutt, 2005). The notion of performance plays a central role in the use of the LFA – it is seen as 'providing a clear set of expectations of what the work should achieve, which makes it easier for people to make the case for funding and also sets up a framework for reporting on progress. This is appealing to both donors and NGOs, who have to justify their decisions and spending in terms of delivery.' One NGO stated in a 2005 Sida-sponsored study of the use of LFA that the main reason for applying LFA principles in programming is:

"...the strongly felt need to be able to present, much more than before, results of the activities we support and implement, that is to demonstrate that our activities really make a difference to the goals and objectives we formulate..."

Despite this stated goal, and the close association between LFA and results-based approaches, it might be expected that LFA would play a prominent role in monitoring and evaluation. However, few agencies routinely compare the progress against the logframe: "for many, the LFA is put to rest once the project or programme begins." Where it is used, the focus is more often than not on 'proving' the logframe – i.e. to look at the expected achievements laid out in the matrix – rather than the work itself. "The logical framework only becomes important again when the headquarters or the donor wants to see it. What is on paper is less important at local level until someone higher up the chain requires the paperwork." As a result of this upward accountability focus, the log frame can become a 'lock frame', making implementing agencies less likely to deviate from the plan even when the context changes.

The planning phase can involve establishing principles to guide the response. This is especially important in complex emergencies, where the context changes so rapidly that it is necessary to be flexible about goals, and to have some frame of reference that can remain relevant in a fluid environment. The basic principles of humanity, neutrality and impartiality are often supplemented by additional principles of accountability, 'do no harm', participation of affected communities, and respect for culture and custom. These principles can be agreed in a specific context, and should be discussed, so that they help to guide the response and provide a framework for establishing priorities, solving problems and making decisions.

An example of a collective approach to establishing shared principles is the Joint Policy of Operations (JPO) drawn up for NGOs working in Monrovia in 1996, based on a mixture of Red Cross principles and the 'do no harm' approach (For more on the use of such principled approaches, see Leader, 2000). By creating a set of shared principles according to which the relief would be delivered, the organisations created a form of humanitarian conditionality for use with belligerent parties.

The planning of relief projects and programmes tends to be closely linked to funding applications and agency concerns about how to maximise impact with available resources. This raises the question of whether planning decisions are made according to where results can best be demonstrated, in line with NPM approaches (Macrae et al., 2002) as discussed above in Section 1.1.3. Whatever the answer, it is evident that agency planning decisions are strongly influenced by anticipated donor response:

'Given the tendency of contract-based relationships to be evaluated against contracted input and output rather than actual outcomes, there is a danger of circularity – problems are 'constructed' and 'solved' in ways that may bear little relation to actual needs' (Darcy and Hoffman, 2003).

This circularity may limit the scope for performance-oriented planning, according to the definition of performance highlighted above (section 1.2.2). Consequently, there is a need for both independent verification of needs and affirmation of related plans. Such mechanisms are thin on the ground, and as a result, planning does not have the central role in establishing the groundwork for humanitarian performance that perhaps it should.

One way to achieve this would be through collective analytical, planning and fundraising tools, which would in principle minimise the potential for donor or agency interests to prevail. There have been attempts to establish mechanisms for this – such as the CAP and the CHAP, facilitated by OCHA – but neither tool has been entirely successful. The CAP, a collective appeals tool, tends to present information about humanitarian need 'exclusively in terms of unmet funding requirements', and is not backed up with accurate field-based indicators of humanitarian needs (Porter, 2002). By contrast, the CHAP, a collaborative planning tool, has only partial buy-in: many NGOs are not active participants, and donors still channel a large part of their resources through other mechanisms or directly to NGOs.

Monitoring: theory and practice

As with planning processes, evidence-based monitoring of ongoing projects and programmes has in theory a central role to play in tracking, assessing and improving performance. The reality is that there are many, varied, partial approaches – old and new – all of which coexist with little harmonisation. These multiple, fragmented systems are overburdened with continuous and intensive requests by stakeholders outside the operational context, for information, reporting and accounting purposes. Donors in particular require considerable reporting detail, often in relation to the objectives set out in funding applications, and in specific formats. This raises questions about the performance implications of setting objectives at the funding stage: how they are set will largely determine the agency's 'success' in achieving them.

Humanitarian agencies seeking to maximise their funding, and reduce dependence on a particular donor, will frequently have to report to ten or more different donors – a very heavy burden for many agencies. The lack of commonality in reporting requirements is a frequent cause of complaint. Some donors require whole-project reporting while others require reporting on only the use of their funds. One donor organisation has been described as demanding so much detailed reporting that several staff members in a funded organisation were dedicated to meeting that donor's information requirements. Most humanitarian organisations also have internal reporting requirements, often different again, and agency membership of particular groupings or federations, or participation in specific projects or initiatives, all carry their own information demands. All of this contributes to significant 'reporting overload' for programme and project staff.

The mass of information derived from these considerable efforts may meet the letter of accountability requirements, but does it contribute to any useful learning in terms of performance? A large – perhaps even the majority – portion of programme-level monitoring effort is directed to meeting external demands. Many of these related to authorisation and budgetary control. This is arguably at the expense of monitoring for learning and improvement. This also limits the value of collecting and collating existing monitoring information in any systematic manner. While there are increasing calls for harmonisation, and for decreased reporting activity, there is a clear tension between achieving this and still maintaining accountability for public funds.

There have been some efforts to coordinate reporting requirements placed on operational agencies, although these have been very limited. Some donors are consciously trying to avoid structuring their requirements in order to preserve a relationship of trust with their partner agencies – but these donors are in a very small minority. To make further progress towards harmonisation, donors need to strengthen capacity to interpret the results of strategic processes so that they become tools for improvement and learning, rather than simply bureaucratic requirements or a justification for annual budget determinations.

Overall, there is a need for monitoring to move from an extractive exercise, focused on providing information for donors and senior management, towards a partnership model in which monitoring is a joint activity aimed at testing and reviewing progress. Tools such as real-time evaluation and outcome and impact monitoring may have the

potential to play such a role, but their use at present is relatively small-scale. Worryingly, all too often such tools are seen as performing an audit function by operational staff, with little scope for management learning.

Evaluation and performance

Evaluation is perhaps the most widely used performance tool within humanitarian agencies, its use having grown considerably since the response to the 1994 Rwanda genocide. Defined as systematic and impartial examinations of humanitarian action, evaluations are intended to draw lessons to improve policy and practice, enhance accountability, and *to assess performance* in relation to policy and/or practice against recognised criteria. A combination of pressures on humanitarian agencies to demonstrate their effectiveness, the availability of guidance on evaluating humanitarian assistance, and ALNAP's encouragement and facilitation of sharing experience on evaluation processes have contributed to the rapid increase in humanitarian evaluation.

There has also been an increase in the number of types of evaluation in the humanitarian sector. The traditional single-agency evaluation after the end of an emergency ('ex post') has now been complemented by joint evaluations and real-time evaluations (during an emergency or intervention). Evaluation can be tailored to specific areas, programme or themes, and can be applied at any scale – they can be used to fulfil performance needs of any depth of span. Evaluation is currently expected to fulfil many of the learning and accountability needs of many humanitarian agencies. However, there is widespread dissatisfaction with evaluations as a tool for improving performance. All too often, they are used for symbolic purposes to fulfil the heavy accountability demands of donors, or to deflect criticisms (Sandison, 2006).

Partly in response to this concern, there is increasing interest in evaluating impact in terms of the ultimate effects on affected people's lives. This brings a degree of objectivity to evaluative efforts which would, it is argued, make them harder to deflect. However, undertaking such assessments in relief settings involves a number of considerable challenges. Impact assessment is explored in detail in a parallel study, presented as Chapter 2 of this RHA, which highlights the importance of clarifying potential impacts as part of planning processes.

Frameworks and tools

There are several frameworks and tools useful in programme and project performance. The Sphere project aims to develop standards for humanitarian NGO performance. It includes the Humanitarian Charter, which attempts to establish a legal framework for the humanitarian community, and Minimum Standards which aim to quantify, in selected areas of work, what it takes to satisfy the legal obligations laid out in the Charter. Key indicators for the fulfilment of standards are given for water supply and sanitation, nutrition, food aid, shelter and site planning, and health services. Agencies' ability to achieve the Minimum Standards depends on many factors, some of which may be beyond their control, such as whether they can operate in conditions of reasonable security. Nevertheless, Sphere offers programme managers a benchmark against which the performance of the core, life-saving sectors may be established and compared. Sphere is frequently used to help plan, design, monitor, evaluate and coordinate emergency relief efforts. Perhaps more importantly, it has been useful in developing debate on humanitarian performance in field settings.

The OECD-DAC evaluation criteria, as adapted by ALNAP, are widely used in humanitarian evaluation, and correspond closely to the 'three Es' approach (outlined above in Box 1.5) – with ideas of economy, efficiency and effectiveness adapted for use in emergency operations. The Red Cross Code of Conduct (see Box 1.6 above), is a set of principle-based commitments, and has been used to guide evaluations of the collective performance of member agencies of the Disasters Emergency Committee (DEC) in the UK. Evaluation teams adopting this approach have found that the Code can be used to measure quality in evaluations and to strengthen accountability, and recommend that the sector develops indicators for compliance with the Code (Vaux et al., 2001).

While both the OECD-DAC criteria and the Red Cross Code have seen applications in evaluative efforts, their scope for performance approaches are wider. For example, both could play a role in performance frameworks which inform planning and monitoring. Some organisations do attempt to specify performance criteria which include factors corresponding to OECD-DAC criteria, such as timeliness, cost-effectiveness, participation and so on. The use of common criteria for both setting performance goals and subsequent monitoring and evaluation has the potential to facilitate more direct and explicit links between strategy, reporting and evaluation functions (ALNAP, 2001).

One way in which this is achieved is through integrated planning, monitoring and evaluation systems. Of particular note, and perhaps underutilised at present, is the Quality Compass system: a voluntary, open-source method of quality assurance for humanitarian assistance. It was developed by Groupe URD with the specific intention of making explicit the linkages between planning, reporting and evaluation. Critical of prescribed standards of performance (such as Sphere), Groupe URD has argued for a 'shift from quality control – ex-post evaluation and verification of compliance to standards – to quality assurance – prevention by the management of critical points during the project cycle.' (BOND, 2006). The Compass method, featuring 12 criteria arranged in a compass rose, combines both project management and project evaluation to promote continuous quality improvement. Of particular note is the fact that the COMPAS puts notions of beneficiary accountability at its heart. Such principles are also at the heart of the complaints and feedback mechanisms set up by HAP members, leading to a number of mechanisms being used in different projects and programmes.

1.2.5 Understanding performance at operation-wide levels

Development agencies wanting an operation-wide perspective on performance tend to use country-level assistance strategies, targets and evaluations. The MDGs also provide an important benchmark for the overall contribution of development aid in a particular country or region.

Within humanitarian relief, some agencies do undertake analysis at a country level, which usually requires aggregation of performance information about projects and programmes. However, this does not give a picture of operation-wide performance. For this, performance needs to be aggregated across organisations, looking at both the performance of delivery in specific areas (health, water and sanitation, etc) as well as common functions which need to be established across organisations.

Four mechanisms are of particular interest here – joint evaluations, tracking and surveillance systems, beneficiary-survey tools, and coordination bodies, as discussed below.

Joint evaluations

As at programme level, joint evaluations are a key mechanism for performance assessment at the operation scale. The Joint Evaluation of Emergency Assistance to Rwanda (JEEAR) in the mid-1990s was the first system-wide joint evaluation of an international humanitarian response. Despite the JEEAR's success in revealing performance issues that would probably not otherwise have been detected, and in raising important humanitarian policy issues, it was ten years before the next comparable joint evaluation – by the Tsunami Evaluation Coalition (TEC) (Telford and Cosgrave, 2006). The recent ALNAP review of joint evaluations (Beck and Buchanan-Smith, 2008) found joint evaluations to be of better quality than sole-agency evaluations, that they offered better coverage of key issues, and they were generally underused. At the present time, there are more joint evaluations which bring together 'clubs' of similar agencies – for example, health evaluations undertaken by clubs of UN agencies, or cross-NGO evaluations undertaken by ECB agencies.

That review recommended that a third joint evaluation should be undertaken across the humanitarian system. Undertaking system-wide evaluations more regularly, perhaps every two years, would have several benefits. They could be approached more systematically, and the governance, management and funding arrangements could be established and would not have to be recreated each time. This would lead to substantial reductions in transaction costs and set-up times, compared with those of one-off endeavours like JEEAR and TEC. Evaluation methods and protocols could also be put in place and refined after each successive evaluation. Moreover, it would be easier to assess the extent to which the lessons from previous system-wide evaluations had been implemented in the next operation to be evaluated.

Tracking and surveillance techniques

Tracking and surveillance tools provide another potentially important mechanism for understanding performance at operation-wide levels. Such approaches have tended to focus on indicators of health and nutrition in societies and communities to which aid is delivered. One reason for this is because health indicators reflect biological changes in people, and so can be assessed and understood across populations and contexts. In contrast, other aspects of lives and livelihoods are harder to measure and more context-specific and therefore less comparable.[3]

In most humanitarian operations, retrospective surveys – after an intervention – are the only means by which information on mortality rates can be obtained, usually via questionnaire. In the last few years, retrospective surveys of how many people have died in a preceding period, and likely cause of death, have been undertaken in only five operations – in DRC, Iraq, Afghanistan, Darfur and northern Uganda.[4] Collecting these data on an ongoing basis requires prospective (real-time) surveillance which implies ongoing, systematic recording, analysis and interpretation. There are now several initiatives to track information on health and nutrition at an emergency-wide level.

- SMART has its origins in North American legislation requiring public-sector bodies to demonstrate performance results. It uses the two basic indicators of crude mortality rate and under-five nutritional status.[5] SMART is now established as an interagency initiative, begun in 2002, to improve the monitoring and evaluation of humanitarian assistance interventions by developing standardised methodologies for comparability across emergency settings.

- The Complex Emergency Database (CE-DAT) managed by the Centre for Research on the Epidemiology of Disasters (CRED) was launched in 2003. CE-DAT manages and provides access to a database of health and nutrition surveys undertaken by agencies in different humanitarian operations. Currently CE-DAT contains over 1,800 surveys conducted in complex emergencies in 46 countries since 2000.[6] As more agencies use the SMART methodology, the proportion of directly comparable surveys in the database is increasing.

- The Health and Nutrition Tracking Service (HNTS) aims to strengthen and develop existing national health information systems and to use humanitarian agencies to help fill gaps in capacity. It began working in Uganda in 2008, and is expected to expand operations to other countries.

Most of these tracking systems are oriented to the needs of donors and the UN, and they tend to lack focus on field capacity to strengthen primary data collection. There is minimal or no tracking of health-status outcomes as humanitarian operations unfold (Mock and Garfield, 2007). Also, and perhaps most significantly, they do not overcome the difficulty of attributing observed changes to particular activities or interventions. As leading analyst, John Seaman has argued in the case of nutritional tracking: 'it is only rarely possible to establish the relationship between the actions

of the international system and the nutritional result in other than circumstantial terms, and virtually impossible to establish clearly the failure of any specific agency' (Seaman, 1999).

Most of these initiatives focus on gathering indicators with little contextual information, little explanation and little effort to identify the meaning of the information gathered. Agency managers who have used these systems have indicated that the data collected are too small-scale and insufficiently detailed to be useful for decision making: 'as a result, there is a disconnect between the supply of information and what the decision makers at HQ and the field actually need… we are frustrated with the current situation. We will make decisions with or without needed information, and most frequently it is without' (Mock and Garfield, 2007).

Beneficiary surveys and assessments

As an alternative to monitoring health status and other indicators to measure operation-wide performance, outcomes and impact can be assessed in terms of the views of programme beneficiaries and the affected population.

Despite the potential of such mechanisms, it is clear that beneficiaries are still only typically involved at the needs assessment stages of emergencies, and at these stages, they feel 'over-assessed and under-consulted'. One common argument against the use of beneficiary feedback put forward by implementing agencies is that end-users of humanitarian services are not seen to be in a position to put forward balanced views on the performance of operational responses. While this concern may be well-placed, there are mechanisms to address such biases, and it is more likely that underlying this aversion is the fear of negative consequences for the profile and branding of agencies. This lack of beneficiary feedback is in sharp contrast to many other services provided in private and public sector work.

Perhaps because of this lack of trust and associated fears, while the principle of consulting beneficiaries has been widely supported throughout the humanitarian system for some time, such consultation has only begun to happen more systematically in the last few years.

Interestingly, academics and independent institutions have carried out more of this kind of opinion-based assessment of performance, working across entire

operations, than has been done within specific agencies or projects. Two recent examples stand out:

• the Fritz Institute, following the Indian Ocean tsunami, has demonstrated the feasibility and value of independently conducted structured questionnaire surveys of large samples of beneficiaries

• the Listening Project is felt to have demonstrated the value of open-ended conversations with beneficiaries, highlighting the potential benefit of complementing questionnaire-based surveys with group discussions and a more ethnographic approach.

There appears to be scope to strengthen such mechanisms through improved sharing of the range of tools used, and developing databases to make the results of such surveys available, in a similar way to CRED.

Coordination mechanisms and their role in operation-wide performance

Because of the need to aggregate data and information, coordination bodies have a special role in operation-wide performance. In recent years, OCHA has been increasingly effective in establishing systems for agencies to report on their activities and outputs on a regular basis across an entire operation. In most operations many agencies now submit monthly reports to their nearest OCHA office and the information is collated and shared with agencies and donors. As part of this sharing and to make the information available to a wide audience, OCHA has established the 'Who does What Where' (3W) website and database.[7] It is unknown, however, what proportion of agencies within each operation actually participates in the OCHA 3W reporting.

This system organises information by use of the Clusters, an approach introduced in 2005 to strengthen partnerships in key sectors of humanitarian response. By formalising the lead role of particular agencies in each sector, and giving that organisation the responsibility for coordination and as a provider of last resort, the clusters aim to work to improve predictability, response capacity, accountability and coordination. Although controversial and decried by some parts of the humanitarian community, there are now discussions underway to strengthen the role of the clusters in assessing and tracking performance.

Another related initiative is the Assessment and Classification of Emergencies (ACE) Project, launched in October 2007. The description of the ACE project within the Emergency Response Coordination section of OCHA in 2008 states: "The first task towards developing a common approach will be to select and pilot the use of a limited number of humanitarian indicators for each cluster or sector, capitalizing on work previously completed by the agencies. The agreed indicators and definitions will then be used to develop and pilot a common humanitarian classification system in two countries in 2008, building on the experience of the IPC. The aim is to test the potential for expanding the IPC beyond its main focus on food security, to serve as the classification system for the wider humanitarian community."

A main conclusion of the work highlights the struggle to link information throughout the "assessment continuum", which covers pre-crisis vulnerability analysis, preparedness (including contingency planning), post-crisis assessment (both immediate and longer-term) and performance monitoring. The ACE project will help to facilitate a shared common, shared understanding of these linkages and to improve comparability.[8]

Mechanisms of NGO coordination, such as the NGO Coordination Committee in Iraq (NCCI) are useful for information sharing, but also for programmatic and policy-oriented collaboration and joint planning. The NCCI has proved important for monitoring the local situation, maintaining working contacts with local actors, promoting the notions of neutral and impartial humanitarian work, and feeding and facilitating the two-way flow of information about what is being done and what is being achieved. One challenge here, arising partly as a consequence of successful coordination, is that the resources available for such mechanisms do not always keep pace with the expansion of the membership and the growing number and complexity of demands (NCCI).

1.2.6 Performance at organisational and cross-organisational levels

Most development and humanitarian organisations of any size, or with a strong commitment to manage the organisation effectively, have developed a number of internal methods and systems for measuring their performance. These might be driven by public pressures, statutory requirements or voluntary development.

The performance approaches described above (in Section 1.2.4) sometimes fit within such mechanisms or, more commonly, operate alongside them. Most humanitarian agencies have some form of strategy or policy relating to their emergency work, and which articulates mission and priority areas. Such statements also provide a framework for gathering information and reporting on performance, and can be divided into sub-categories, often in a hierarchy of objectives. An example is shown in Box 1.8.

Data for such assessments should ideally come from performance information systems that monitor key performance indicators (KPIs) in relation to the strategic

---

Box 1.8	WFP's Strategic Objectives and related performance measures

For the period 2006–09 WFP's Strategic Objectives were:

1 save lives in crisis situations

2 protect livelihoods in crisis situations and enhance resilience to shocks

3 support the improved nutrition and health status of children, mothers and other vulnerable people

4 support access to education and reduce gender disparity in access to education and skills training

5 strengthen the capacities of countries and regions to establish and manage food-assistance and hunger-reduction programmes.

Each of these objectives has associated statements of output, outcome and indicators. For example, for Strategic Objective 1:

■ the output statement is 'Timely provision of food in sufficient quantity for targeted beneficiaries in conflict and disaster affected areas'

■ the outcome statement is 'Reduced and/or stabilised acute malnutrition in an identified population in conflict and disaster affected areas'

■ Indicators for this outcome are the prevalence of acute malnutrition among under-5 children, mortality rate and under-5 mortality rate.

Annual performance reports then analyse actual performance against the Strategic Objectives using standardised project reports prepared by every country office and regional bureau.

objectives of the organisation. These are invariably considered in relation to single humanitarian organisations, and are less commonly used across whole sectors or operations. This provides an interesting comparison with the development sector's use of the MDGs, and efforts in the wider public sector such as WHO's health systems indicators. Information may also be derived from project and programme monitoring systems established within agencies, evaluation reports, and international shared sources such as tracking systems. Different organisations have developed KPIs on particular themes. For example, IFRC has spent some time developing and testing a set of KPIs for logistics, and People In Aid assesses and improves human-resources performance using a code-and-compliance mechanism.

Experiences with organisational performance management

How are organisational systems doing in practice? Two UN RBM mechanisms were recently evaluated by the UN's OIOS, both of which provide useful lessons. Effective results-based management requires that each programmatic activity contributes to the achievement of the strategic objectives. At WFP this is ensured through a performance tool called the Indicator Compendium, setting out all the outputs and outcomes that WFP needs to deliver in order to achieve its strategic objectives (see Box 1.8), and the indicators it will use to measure its performance against these. The Compendium also describes the expected results that WFP intends to achieve to help it deliver against its management objectives, and the indicators that it will use to measure its achievement of the expected results.

The findings of a recent review (NAO, 2008) of WFP's RBM were three-fold. First, while it was acknowledged that WFP's strategic objectives represented valuable goals, their aspirational nature meant that WFP was unable to measure its performance against them. None of the strategic objectives are specific, and they could not be quantified. Second, the performance indicators do not measure all the factors required to achieve each output or outcome. Therefore, achievement against all the indicators listed for a particular output or outcome in the Indicator Compendium would not necessarily ensure achievement of the associated outcome or output. Third, learning from the results of WFP's performance applies mainly at country and regional levels, but objectives, outcomes, outputs and indicators are not regularly or routinely revisited in the light of actual performance. Moreover, there is no systematic procedure for learning from performance systems across different projects, countries and regions.

The experience of OCHA with RBM has also involved considerable difficulties, including in collecting data necessary to assess performance against agreed indicators. The performance-monitoring process was seen as immaterial and of little value for informing decision-making. Performance measures did not greatly influence resource allocation, or work programmes. Overall, programmatic performance assessments were perceived as compliance requirements or a 'paper exercise', rather than an opportunity to improve performance (OIOS, 2007).

Perhaps the most extensive performance approach in the Red Cross is the Planning for Results (PfR) system introduced at the ICRC in 1998. PfR is a good example of a 'balanced' approach in the humanitarian system, and covers planning, budget construction and appeals, implementation, logistics, financial and human-resource management and evaluation. It is an integrated system used at every level of the institution, and at all phases of the project cycle. The objectives of the system are to enhance the performance of ICRC's operations, to promote a results-oriented management culture and to evaluate results and impacts. The collection of indicators is a central dimension of PfR. Indicators are defined in relation to outputs, outcomes and, whenever possible, the impact of programmes. PfR was highlighted by DFID's MEFF assessment as a highly effective performance system (DFID, 2004)

ActionAid began using the Accountability, Learning and Planning System (ALPS) in 2000 in order to: reduce the amount of work involved in programme reporting, reduce reliance on written reports and learn more from existing programmes. A key component of ALPS is the replacement of country reports with annual participatory reviews and reflections:

> 'By reducing the drudgery of written reporting… ALPS should make space for staff to interact more with partners and poor people. And by introducing processes of review and reflection, it is intended to help poor people, our partners and ourselves, to learn from our experiences and those of others in order to continuously improve the quality of our work' (ActionAid, 2000, p. 6).

Since its adoption, ALPS has apparently been internalised in the organisation and resulted in explicit reflection-learning-action cycles, increased downward accountability, a culture of transparency and better understanding of impact. Monitoring now has more of a participatory than a reporting focus (David and Mancini, 2003).

The Humanitarian Accountability Partnership (HAP) has developed a quality-assurance tool for humanitarian non-governmental agencies to assess and re-orient their performance and quality-management systems more towards their intended beneficiaries, rather than to managers and donors. The Humanitarian Accountability and Quality Management Standard published in 2007 consists of a set of accountability principles and benchmarks, with an accompanying certification and accreditation system. HAP's tools provide a solid basis for agencies seeking to improve the quality and accountability dimensions of their work across the organisation.

The relationship between performance and evaluation

While the relationship between performance and evaluation is complex, there are several key distinctions between performance systems and systems for evaluation and learning.

- Performance systems tend to rely on self-evaluations by project or programme managers, while evaluations are usually conducted by external teams.

- Performance monitoring is often mandatory, especially for larger projects, while evaluations are discretionary and optional.

- Performance reports involve straightforward presentations of data in standard, comparable formats, enabling comparability but not real depth of analysis, while evaluations are highly customised and focus on fewer issues in greater depth.

- Performance monitoring determines whether results were achieved, while evaluations explain why and how they were achieved, or not, and seek to analyse and understand the programme or project context and the factors influencing performance.

- While performance tends to emphasise quantitative indicators, evaluation uses blended approaches that are better for understanding context and generating recommendations.

The rise of humanitarian performance monitoring has led to some confusion about the function of evaluations. Although in principle the two approaches should be complementary, the differences and distinctions are not always clear-cut. There are

several ways in which the two approaches could be integrated. However, in practice, they are often in competition for the same resources – money, staff time and others. In some agencies, such as USAID, a major commitment to performance management has coincided with a corresponding reduction in the scope and number of evaluations, suggesting a trade-off between the two functions.

Cross-organisational approaches to performance

There are several approaches to performance management shared by different 'clubs' of organisations, such as NGOs, donors, the Red Cross Movement or UN organisations. There is none that involves the entire humanitarian sector, with the only shared perspective deriving from common humanitarian principles and legal frameworks. Given the voluntary nature of such efforts, cross-organisational

Box 1.9	The DEC accountability framework
Accountability priority	**Definition**
We run well-managed appeals	• Efficient and effective fundraising
	• Timely and appropriate allocation
	• Review of performance
	• Accountability to stakeholders
We use funds as stated	• Ensuring sound financial management at agency and partner levels
We achieve intended programme objectives and outcomes	• Maximising the potential for programmes to achieve objectives and outcomes which respond to a demonstrated need
We are committed to agreed humanitarian principles, standards and behaviours	• Fulfilling the principles embodied in the Red Cross Code of Conduct, Sphere, People in Aid
We are accountable to beneficiaries	• Taking account of, giving an account to and being held to account by disaster survivors
We learn from our experience	• Improving performance based on lessons learnt

initiatives are often criticised for their lack of compulsion or sanctions for non-compliance.

Perhaps the most well-known cross-organisational performance framework in the sector is the Red Cross-Red Crescent-NGO Code of Conduct, discussed above (Section 1.2.1). Although the potential for the Code to be used as a framework for performance is clear, and it has been used for evaluative purposes, a review conducted on the tenth anniversary of the Code's launch found that 'little explicit use [is] made of the Code in humanitarian practice' (Hilhorst, 2005). More recently, the DEC's accountability framework has been developed to be implemented across DEC agencies, and brings together a number of different components of performance (Box 1.9) as relates to DEC funds.

Work by the ECB initiative, a cross-NGO initiative – has developed guidance on aspects of performance, specifically relating to accountability, and have aimed to improve participating organisations emergency capacities. Across the UN agencies, cross-organisational performance has been the focus of a series of reforms, with significant efforts to improve cross-organisational humanitarian response at global and field levels. These include IASC-endorsed initiatives, such as the cluster leadership approach, and donor-led initiatives, such as pooled funding mechanisms like the CERF and common funds at country level.

These reforms are collectively intended to contribute to five distinct aspects of performance: Needs-based responses; Sufficient capacity to respond; Predictable support for operations; Effectively led & coordinated responses; Transparent & efficient response systems.

One of the challenges facing the reforms is how to establish and demonstrate a basis for improved performance. To date, evaluations of the cluster approach and CERF, and an HPG report, have highlighted the importance of establishing measures to assess the performance of the reforms in appropriate ways. This in turn would require determining key aspects of system-wide performance (HPG website; Barber et al., 2008; Stoddard, 2007).

Another related tool is the Organisational Self Assessment Tool developed by the Humanitarian Futures Programme, which aims to assess the readiness of organisations to become more anticipatory, adaptive and collaborative in the face of future humanitarian challenges. This has been used within UN country teams and a

number of other organisations, and has made an important contribution to discussions of UN performance (HFP, 2007).

The Good Humanitarian Donorship (GHD) initiative aims to promote good practice in donor financing, management and accountability, including enhancing learning and accountability. GHD has also become a cross-organisational forum for donors. By defining principles and standards it provides both a framework to guide official humanitarian aid and a mechanism for encouraging greater donor accountability. Based on the GHD principles, the Humanitarian Response Index (HRI) is:

> 'a tool designed to measure how well humanitarian donors are performing relative to their commitment to the Principles. The HRI is intended to help identify and understand donors' strengths and weaknesses in the area of humanitarian action in order, ultimately, to improve the quality of humanitarian action and alleviate human suffering in crisis situations' (DARA, 2007).

Within the annual report of the HRI, overall results for each donor country are ranked from best to worst. Despite a strong reaction by some donors to the ranking process, the HRI is now approaching its third year.

1.2.7 Approaches to system-wide performance

The discussion initiated in Section 1 above about system-wide performance is particularly pertinent for the humanitarian system. At present, there is hardly any attempt to assess performance of the humanitarian sector as a whole, on either a one-off or regular basis, in order to reveal trends. The last ALNAP RHA highlighted the inability of the humanitarian system to assess its own performance (Mitchell, 2008). There are still no baselines, no agreed definitions of performance and an absence of any kind of mechanism able to track performance. The TEC identified little change since Rwanda – but is the Rwanda crisis the best baseline comparison, or the only baseline that could be used?[9]

The humanitarian system is characterised by a high degree of 'mutual dependence' (Stoddard, 2003): the increasing scale of emergencies and the comparatively limited response capacities mean that even the largest organisation is incapable of

launching an effective response on its own. Despite this, the policy apparatus within the humanitarian sector is not as advanced as in the development sector: there are few if any humanitarian equivalents to the MDGs or the Paris Declaration on aid effectiveness.

Approaches to date (such as the UN reforms) have tended to focus on specific groups of actors, or on specific instruments (such as Sphere), or specific forms of quality and accountability. In contrast, a collective performance framework that could enable preliminary assessment of the 'state of the system', and set a baseline for establishing collective goals and assessing progress against them, could form a vital first step to a more relevant, strategic and credible international humanitarian system. There may be scope to synthesise existing material to provide a fuller picture, which would be a necessary step towards regular system-wide performance assessment.

There have been some landmark reviews of the overall humanitarian system. The *Humanitarian Response Review* was published in 2005 following the 2004 tsunami. It was a sustained reflection on systemic issues, with particular attention to coordination, financing and leadership, which gave momentum to subsequent UN reforms. The longest-running series of regular reports or surveys of the humanitarian sector is the *World Disasters Report*, published annually since 1993 by the International Federation of Red Cross and Red Crescent Societies. There is a range of other reviews and mechanisms that touch on some aspects of performance, but only some. Many of these reviews vary their themes from year to year, and so do not provide consistent, regular analysis of the same issues. Some of the datasets are weak, and so have limited value for performance assessment.

Existing foundations for overall humanitarian performance assessment

The last 10–15 years have seen the development of a significant capacity for independent and objective research and analysis within the humanitarian sector. The ODI's Humanitarian Policy Group, Tufts University's Feinstein International Center and New York University's Center for International Cooperation produce a high-quality research and analysis on humanitarian issues. Research and analysis capacity on disaster mitigation and humanitarian issues is also being steadily developed in Asia, Latin America and Africa. Research from all these sources often includes analysis of performance aspects of the humanitarian system. Collectively

therefore, this may be seen as not only as another source of information on humanitarian performance, but also as an indirect means of assessing and tracking the performance of the system.
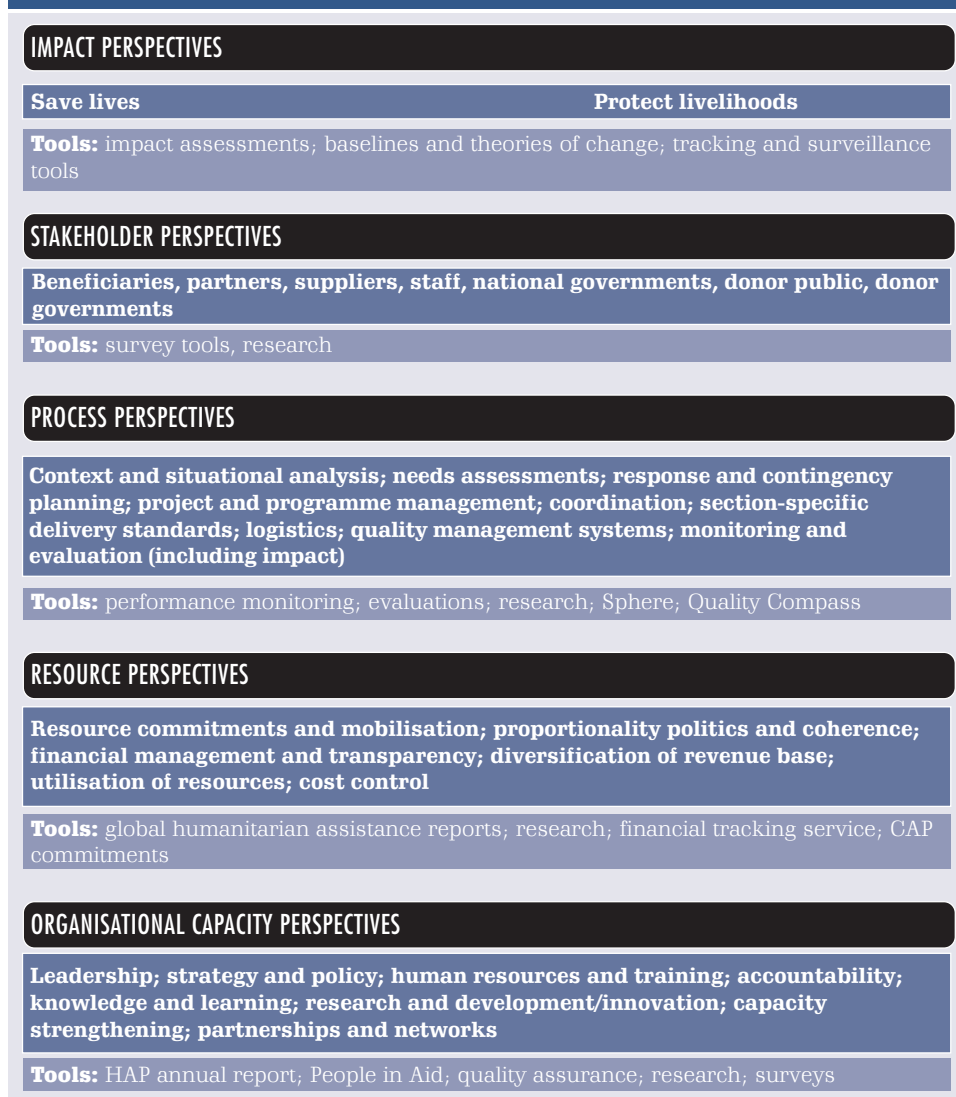
A system-wide process of performance assessment would have several potential benefits, and could, for example:

• synthesise existing material into a comprehensive representation

• integrate different dimensions of performance

• generate indicators of performance assessment

• report on performance against agreed criteria, and on the evidence base for the judgements being made

• provide a platform for debate on performance throughout the system

• allow for adaptations and adjustments, both incremental and radical, to the humanitarian business model

Figure 1.10 is a suggested initial framework for an integrated, system-wide view of humanitarian performance. It incorporates elements of the Balanced Scorecard, Strategy Map and EFQM approaches.This framework is for illustrative purposes, and should also be useful in starting to conceptualise system-wide performance. The next steps in developing this would be adapting, testing through pilot mechanisms and further refining. In particular, the weight given to different components in an overall picture of performance will need to be clarified, and in doing so, different priorities and interests will need to be negotiated. Once agreed, this will need to be revised regularly to ensure the model of the system that it represents is a dynamic one which allows for adaptations and innovations in the structure and function of the system. For example, were radical changes to take place in the sector, changes would need to take place in the performance framework, if not in the range of components then at least in the weighting given them in assessing overall performance.

Considerations of system-wide performance also requires understanding of issues which lie 'beyond the system' but which might influence or be influenced by the

Figure 1.10 An illustrative view of system-wide performance using a balanced approach

IMPACT PERSPECTIVES

Save lives **Protect livelihoods**

Tools: impact assessments; baselines and theories of change; tracking and surveillance tools

STAKEHOLDER PERSPECTIVES

Beneficiaries, partners, suppliers, staff, national governments, donor public, donor governments

Tools: survey tools, research

PROCESS PERSPECTIVES

Context and situational analysis; needs assessments; response and contingency planning; project and programme management; coordination; section-specific delivery standards; logistics; quality management systems; monitoring and evaluation (including impact)

Tools: performance monitoring; evaluations; research; Sphere; Quality Compass

RESOURCE PERSPECTIVES

Resource commitments and mobilisation; proportionality politics and coherence; financial management and transparency; diversification of revenue base; utilisation of resources; cost control

Tools: global humanitarian assistance reports; research; financial tracking service; CAP commitments

ORGANISATIONAL CAPACITY PERSPECTIVES

Leadership; strategy and policy; human resources and training; accountability; knowledge and learning; research and development/innovation; capacity strengthening; partnerships and networks

Tools: HAP annual report; People in Aid; quality assurance; research; surveys

system. Examples drawn from the international sphere might include the role of the military, the private sector, and other external sectors or actors. At the national level, this would include the role of local and national actors, as well as diaspora groups and remittance flows. Such issues are currently explored mainly through research work, but there may be scope for a more systematic consideration to inform regular system-wide performance assessments.

1.3 Summary reflections, conclusions and recommendations

1.3.1 Reflections: key challenges for humanitarian performance

Performance frameworks have been used in a variety of ways within the humanitarian sector, and this use has been influenced by several of the models and approaches reviewed in this study. For example:

- needs assessments are built on the principle of evidence-based practice and policy

- Sphere Minimum Standards are built on ideas of total quality management, while the Charter builds on rights-based approaches

- People In Aid uses a social-audit methodology

- results-based management approaches and ICRC's PfR and use new public management approaches

- Quality Compass uses project-cycle methodology combined with quality assurance techniques

- The OECD-DAC criteria adapt the '3E' model for assessing government performance

- HAP is built on ISO9000 principles and rights-based approaches

- ALPS uses rights-based approaches.

It is useful to revisit the four-part classification of models of performance systems (as discussed above in Section 1.1.6):

- *Traditional/pre-performance models* – no systems for gathering data or sharing it, and performance systems are intuitive and informal

- *Performance administration* – commitment to and expectation of performance assessment and management, but poorly developed and ad hoc, administrative rather than managerial

- *Management of performances* – a number of different systems operate for different, unconnected management functions; diverse range of management of distinct performance areas

- *Performance management* – coherent, integrated, consistent and comprehensive. Different performance assessment systems are used for different purposes, with coordination across the different systems.

On the basis of the above, it would appear that the humanitarian system overall is somewhere between models 2 and 3, and therefore quite a distance from the ideal of model 4. Just a few agencies are well developed in their own efforts. To progress towards model 4, representing an integrated, coherent and comprehensive system, there are a number of key challenges for the humanitarian sector, as detailed below.

The challenge of multiple approaches to performance

The exploratory phase of the Humanitarian Performance Project revealed an extraordinary range of projects, initiatives and approaches in monitoring and reporting on performance within the humanitarian system. There are a lot of different ideas, systems, frameworks and tools, and a lot of data, reports and demands. However, the current humanitarian performance landscape is patchy and erratic – the different initiatives have been established for different purposes and focus on different aspects of the system and different aspects of performance. Many operate in parallel to each other and some overlap.

It does not seem unfair to say that humanitarian performance is a 'mish-mash' of concepts, ideas and tools, many of which are work-in-progress and only loosely connected to each other. The last five years have seen real steps towards information and data sharing to foster an effective and coherent humanitarian response, although there are still some gaps. But the problem is not so much lack of information on humanitarian assistance but *fragmentation* between monitoring and measurement systems, which have often been set up for very different purposes, without common standards, based on different methodologies and categorised according to different definitions (Development Initiatives, 2008).

However, as Darcy and Hoffman (2003) suggest, this may be the only workable model of operation for the humanitarian sector. They identify two models of sector-wide systems (citing Schofield, 2004):

> 'One, the 'systems' model, is highly structured; all agencies cooperate to achieve the common aims... This model borrows from governments and the military, where information is gathered at the base of a hierarchical pyramid, and passed to decision-makers at the top. The second model – the 'service' model – is a much looser arrangement. Here, individual services fill particular niches. Each agency or individual chooses whether and how to use these services. Service providers are effectively competing to achieve the common aims... This model... derives from commercial news and market information services, is closest to information services such as ReliefWeb and OCHA's Integrated Regional Information Networks (IRIN).'

If this is the case, it may not be productive to focus on the fragmented and multiple approaches that exist in the sector. Instead, the performance discussion should focus on how to establish a meaningful synthesis of the range of approaches and methods that are available. This should be done both at the conceptual level and also in terms of the actual data available.

The challenge of comparable assessments of performance

Despite the considerable number of systems gathering data, and a range of initiatives which attempt to address performance issues, most of these initiatives tend not to use the data to reflect systematically on performance in a regular, comparable manner. Many performance-related efforts focus on what should be done, and stop short of reflecting on how well it is actually done. By contrast, performance mechanisms that are promoted hard – especially by donors – are generally not used to their full potential, becoming instead an administrative reporting chore, as with RBM in some UN agencies. Systems established to track key indicators have been developed outside organisations, tend not to include any way of establishing attribution of humanitarian performance to the changes they monitor, and tend not to be used within agencies to inform their analysis.

The assessment and reporting focus in humanitarian work is mainly on needs assessments initially, and then on reporting activities and outputs during operations, with assessment of outcomes and impacts receiving less attention. This 'front

loading' of assessments within operations is often related to the main incentives for agencies to collect information on needs for use in planning their interventions, raising funds and allocating resources between different operations. Even where performance systems are well-developed, performance assessments are seldom used as the basis for resource allocation. The sanction of discontinuing funding for agencies unable to demonstrate positive outcomes appears to be rarely, if ever, deployed. This suggests that the incentives for performance systems that demonstrate outcomes and impacts are considerably weaker.

This is not to say that there are no reflections on how well things are done. Evaluation plays a central role in current humanitarian performance, and is appealing for several reasons: it can be tailored to the needs of the agency concerned, and has potential for generating learning and accountability in some form even where monitoring systems are weak or absent. The likelihood of more joint, system-wide evaluations of selected operations is a welcome development and will provide an important source of information for assessments of the overall performance of the humanitarian system.

However, in the long run, evaluation may not be able to provide a sufficient basis for longitudinal monitoring of trends in performance due to its discrete application and its lack of comparability. Differences between contexts, methods and the composition of evaluation teams make it difficult to compare results directly between evaluations. While it is realistic to expect evaluation to play a major role in performance assessment within the humanitarian system for many years to come, it should no longer be seen as the sector's only or even major performance tool.

The challenge of widening the depth and span of performance approaches

Research for this study has also revealed increasing numbers of performance mechanisms at the operation-wide, organisational and cross-organisational levels, despite the still-significant focus of performance-related effort at the project and programme levels. However, there is still no systematic approach to understanding performance at a system-wide level, with the system instead relying on a mixture of different sources, including reports, research and the (very) occasional joint evaluation.

This finding might be expected, given the nature of the humanitarian system, with many agencies undertaking the implementation of projects and programmes and fewer agencies responsible for coordination and leadership at sector and operation levels. As a result, therefore, most of the focus of performance efforts is at the programme level of individual agencies, while methods for performance assessment at the operation level and above remain to be developed. Efforts such as ACE outlined earlier are important developments in this regard.

There is also a key issue in relation to the span of performance approaches. Most, if not all, seem to stop short of assessing anything beyond outputs. While this is understandable, given the challenges of (a) doing such assessments and (b) incorporating them into a wider performance framework, this necessarily limits the scope of performance approaches in the sector. At best, we can assess internal performance *efficiencies*, and the issues that are faced. A solid and credible understanding of humanitarian aid *effectiveness* may be some way off.

The challenge of a system in flux

One of the most serious concern for any performance mechanism is how it enables the system that is being assessed to adapt and evolve, whilst maintaining a clear sense of consistency. This is a particularly challenging issue for the humanitarian sector, in which some of the most important changes, which would improve performance, would also involve radical changes to the nature of the system. An accompanying ALNAP study in the 8th Review of Humanitarian Action highlights in particular the three-fold role of disaster prevention, local ownership and partnerships and beneficiary participation in bringing about paradigmatic changes and innovations in the system (Ramalingam, Scriven and Foley, 2009). Were such changes to take place, it may call for a significantly revisions to performance frameworks used in the sector. As suggested earlier, this may not affect the components of the framework, but in the weighting given to different aspects of performance, which would need to change as the system, and priorities within it, change.

The key point is that the performance frameworks employed in the sector should have the scope to move beyond 'single loop learning' – which are improvements undertaken in line with existing practices, policies and norms of behaviour. They should also facilitate *double loop learning* – which focuses on creative and inventive solutions to existing problems, which calls for mindsets, capacities and institutional

space to generate and realise new visions for how work could be undertaken. Even more challenging, but equally important, is for performance systems to facilitate *triple-loop learning*, which involves questioning the entire rationale of an organisation, and can lead to innovative transformations in internal structure, culture and practices, as well as in the external context (ibid).

The challenge of outstanding gaps in understanding humanitarian performance

There are significant gaps in how performance has been conceptualised and tracked in the humanitarian sector, and four are outlined here. This is not to say this is a complete list, but these are most evident on the basis of the present analysis.

First, the views and opinions of key stakeholders, especially beneficiaries and affected populations, are underutilised. Other stakeholders, including partners, Western donor publics and national governments, also appear to be routinely left out of performance considerations.

The relative lack of mechanisms within performance approaches for seeking opinions at beneficiary and community levels bears out the frequent criticism of the humanitarian system that it does not make sufficient effort to gather beneficiary views, and pays little regard to such views when they are available. There are positive signs that this relative neglect is being addressed, and the growing number of agencies preparing themselves for HAP certification is encouraging. The use of surveys appears to be growing, while the Listening Project is another important innovation.

Second, there appears to be considerable scope for improving analysis of how context affects and can constrain performance. This would require some methodological development, and testing in a range of operations, but such analysis has the potential to inform operational planning as well as subsequent performance assessments.

Third, there is scope for more consideration of organisational capacities. The ICRC PfR model provides a useful example of a single system which incorporates factors such as planning, resource mobilisation and logistics, but other aspects – such as capacity strengthening, partnerships and innovation – do not seem to play any role

in performance mechanisms. The performance of southern disaster actors, their relative advantages and strengths, is of special interest here.

The fourth and final gap in performance efforts is the lack of focus on impact and outcomes. Although there is some evidence that this is beginning to change (as described in an accompanying study in this years RHA), there is still some way to go before such assessments regularly and sytematically inform understanding of performance.

1.3.2   Conclusions and recommendations

The humanitarian system has many of the right conceptual components for a comprehensive model of performance, despite having fewer of the necessary data. There is scope and space to elevate performance to a sector-wide concern, mirroring the efforts of the Millennium Development Campaign and the aid-effectiveness agenda. The policy apparatus is much less evolved in the humanitarian than in the development community, but there is still scope for the humanitarian system to start regular assessments of the sector overall, acknowledging problems of definition, measurement and expectations.

Such an agenda for moving towards performance approaches would reflect the fundamental interdependence of the modern humanitarian system. Further, it would contribute significantly towards achieving four key goals of the humanitarian sector.

A performance focus could help the sector to progress: beyond single-agency and towards collective views; beyond outputs towards impacts; beyond internal towards external perspectives, especially those of affected communities and countries; and beyond data for needs and fundraising towards data for performance and accountability.

For those in the humanitarian system who are interested in shaping and implementing such an agenda, four recommendations are offered, as follows.

1   Develop an integrated framework for performance, onto which existing efforts and initiatives can be mapped. A balanced, comprehensive and coherent

framework for humanitarian performance could be discussed and agreed across agencies (an illustrative example developed over the course of the present study was presented in Figure 1.4 above).

2 Explore the potential for using such a framework to address key issues and gaps within the humanitarian system, such as by:

- establishing clearer links between monitoring and evaluation, strategy and other aspects, using existing common frameworks

- improving harmonisation of existing processes for performance reporting, thereby reducing the administrative burden of existing reporting systems

- prioritising gaps to be filled by sustained, sector-wide R&D efforts.

The study presented in this chapter indicates that these gaps include beneficiary perspectives, context analysis, aspects of organisational capacity (including capacity strengthening, partnerships and innovations), and impact.[10]

3 Encourage the selection of common case-study or pilot countries within current and planned projects, and work to synthesise across these initiatives to generate a series of integrated and comprehensive views of performance in the selected settings.

4 Undertake a regular synthesis of system-wide performance, which contains both a reflection of the performance, and the basis on which this reflection is made. This could include:

- rolling out a pilot assessment looking at performance using only existing data, but supplemented with tools such as interviews and surveys, taking account of the existing data gaps

- using this report to initiate dialogue and debate across the sector, and then revising the methodology accordingly

- engaging with key initiatives, suggesting orientation towards the eventual 'synthesised' use of their data outputs in a system-wide performance report

- considering the scope for using such a report to set shared, system-wide goals and to track progress towards them.

- Ensure scope for adaptation and revision is built in at the outset

We would conclude that performance can provide an overarching framework for understanding the vast array of humanitarian initiatives, approaches and techniques. The challenge is to reframe all these initiatives in terms of their separate but complementary contributions to humanitarian aid effectiveness. In moving to improve performance, however, it is important to avoid entrenching the power differences and inefficiencies of the system. Humanitarians should be driven by shared objectives, rather than driven apart by interest, ideology or invective.

It is also essential not to oversell any performance system, but to keep it focused clearly on users and uses, and on realistic expectations for the system within given operating contexts. The approaches should also not promote conservatism and consolidation at the expense of creativity and innovation.

The goal should be to move towards a better understanding of humanitarian effectiveness, while working hard to harness the benefits and avoid the perversities of the performance agenda.

Humanitarian agencies embarking on this journey would do well to remember the sign that Albert Einstein is said to have kept on his office wall, which sums up the required mindset succinctly: 'not everything that can be counted counts, and not everything that counts can be counted'.

Notes

1 http://www.civicus.org/new/media/ putting_beneficiaries_first.pdf

2 Reviewed in a previous ALNAP study on organisational change

3 Mock et al, http://pdm.medicine.wisc. edu/22-5%20PDFs/mock.pdf

4 Les Roberts, Francesco Checchi and Richard Garfield, personal communications, 2008.

5 http://www.smartindicators.org/ SMART_Brief-June_28_2004.pdf

6 The database is publicly accessible and searchable online at www.emdat.be/

7 http://3w.unocha.org/WhoWhatWhere/

8 OCHA in 2008 http://ochaonline.un.org/ ocha2008/htm Sections on Introduction and Financial Position, Policy Development and Studies Projects, Emergency Response Coordination. Final Note for the Record: Informal Meeting on Assessment and Classification Initiatives, 30th January 2008, Geneva.

9 Tong, personal communication.

10 Innovations and impact assessment are discussed in the other chapters in this review, and beneficiary surveys will be considered in a forthcoming study.

References

ActionAid 2000 *Accountability Learning and Planning System*. ActionAid, London.

ALNAP (2001) 10th Meeting Proceedings

Bakewell, O, and A Garbutt (2005) *The Use And Abuse Of The Logical Framework Approach*, Stockholm: Sida.

Barber, M, A Bhattarcharjee, RM Lossio and L Sida (2008) "Central Emergency Response Fund (CERF): Two Year Evaluation," Report submitted to the Under-Secretary-General for Humanitarian Affairs and Emergency Relief Coordinator July 2008, New York: OCHA.

Beck, T, and M Buchanan-Smith (2008) 'Joint Evaluations Coming of Age? The Quality and Future Scope of Joint Evaluations' in *ALNAP Seventh Review of Humanitarian Action*. London: Overseas Development Institute.

Bond (2006) *A BOND Approach to Quality in Non-Governmental Organisations: Putting Beneficiaries First*, London: BOND

Borton, J (2008a) *Inventory of the Principal Projects, Initiatives and Approaches which are Relevant to Overall Performance Assessment of the Humanitarian System*. London: ALNAP/ODI.

Borton, J (2008b) *Summary of the Emerging Results from the ALNAP Humanitarian Performance Project: Exploratory Phase*. London: ALNAP/ODI.

Bouckaert, G, and J Halligan (2008) *Managing Performance: International Comparisons*, Routledge.

Bourgon, J (2008) *Performance Management: It's the Results that Count, Paper for workshop on Government Performance and Results Management*, Presented at APEC, Taipei, 27–28 March 2008.

Brett, EA (1993) 'Voluntary agencies as development organizations: theorising the problem of efficiency and accountability' in *Development and Change*, Vol. 24, pp269–303.

de Bruijn, H (2007) *Managing Performance in the Public Sector* (2nd edition). Abingdon: Routledge.

Burnes, B (2004) *Managing Change: A strategic approach to organisational dynamics*, Financial Times/ Prentice Hall, London.

Clarke, P, and B Ramalingam (2008) 'Organisational Change in the Humanitarian Sector' in *ALNAP Seventh Review of Humanitarian Action*. London: Overseas Development Institute.

Cosgrave, J, and M Herson (2008) 'Perceptions of Crisis and Response: A Synthesis of Evaluations of the Response to the 2005 Pakistan Earthquake' in *ALNAP Seventh Review of Humanitarian Action*. London: Overseas Development Institute.

DARA (Development Assistance Research Associates) (2007) *Humanitarian Response Index 2007: Measuring Commitment to Best Practice.* Basingstoke: Palgrave Macmillan.

Darcy, J, and C-A Hoffman (2003) *According to Need? Needs Assessment and Decision-making in the Humanitarian Sector*, HPG Report 15. London: Overseas Development Institute.

David, R, and A Mancini (2003) *Transforming development practice – the journey in the quest to develop planning, monitoring and evaluation systems that facilitate (rather than hinder) development.* ActionAid paper for DAC conference March 2003.

Development Initiatives (2007) 'Humanitarian Data Access and Transparency Programme (including Global Humanitarian Assistance reports)'. Draft proposal dated 21 November.

Development Initiatives (2008) *Global Humanitarian Assistance 2007/2008.* Wells: Development Initiatives.

DFID (2004) International Committee of the Red Cross (ICRC) Development Effectiveness Summary.

Drucker, PF (1955) *The Practice of Management*. Butterworth-Heinemann, London.

Ebrahim, A (2003) *Accountability in Practice: Mechanisms for NGOs*. World Development, No. 31, 5: 813–829.

ECHO (2004) 'ECHO Strategy 2005: Global Humanitarian Needs Assessment (GNA) and Forgotten Crisis Assessment (FCA); Methodological Notes'. Brussels: ECHO 4/PB D (2004).

Edwards, M, & D Hulme (1995) NGO performance and accountability: Introduction and overview. In M. Edwards and D. Hulme (Eds.) *Non-governmental organisations: Performance and accountability*, 3–16. London: Save the Children.

GHD (Good Humanitarian Donorship) (2003) *Principles and Good Practice of Humanitarian Donorship*. Stockholm, 17 June (www.goodhumanitariandonorship .org).

Graves, S, and V Wheeler (2006) *Good Humanitarian Donorship: Overcoming Obstacles to Improved Collective Performance*. HPG London: ODI.

Griekspoor, A (2001) *Enhancing the Quality of Humanitarian Assistance: Taking Stock and Future Initiatives, Prehospital and Disaster Medicine*. October–December 2001.

Hallam, A (1998) *Evaluating Humanitarian Assistance Programmes in Complex Emergencies*, RRN Good Practice Review 7, September. London: Overseas Development Institute.

Hailey, J, and M Sorgenfrei (2004) *Measuring Success: Issues in Performance Measurement*, INTRAC Occasional Paper Series no 44, INTRAC, Oxford.

Hilhorst, D (2005) 'Dead or alive? Ten years of the Code of Conduct for Disaster Relief', Humanitarian Practice Network Humanitarian Exchange, 29 March. London: Overseas Development Institute.

HPG (2003) *Humanitarian action in conflict: implementing a political economy approach.*

Hudson, M, A Smart and M Bourne (2007) *Theory and practice in SME performance measurement systems*, International journal of operations and production management, Vol. 21 No.8, pp 1,096–1,115.

I&DEA (2006) 'Review of Performance Improvement Models and Tools' Improvement and Development Agency.

IASC (2007) 'Needs Analysis Framework: Strengthening the Analysis and Presentation of Humanitarian Needs in the CAP'. Geneva: IASC CAP Sub-working Group.

IASC/WHO (2005) 'Report of a Workshop on Tracking Health Performance and Humanitarian Outcomes', 1–2 December, Geneva.

JEEAR (Joint Evaluation of Emergency Assistance to Rwanda) (1996) *The International Response to Conflict and Genocide: Lessons from the Rwanda Experience.* Copenhagen: Royal Danish Ministry of Foreign Affairs.

Kanigel, R (1997) *The One Best Way: Frederick Winslow Taylor and the Enigma of Efficiency*. New York: Viking

Kaplan, R, and D Norton (1992) 'The Balanced Scorecard – Measures that Drive Performance' in *Harvard Business Review*, Jan–Feb, pp 74–79.

Kaplan, RS, and DP Norton (2000) 'Having Trouble with your Strategy: Then Map it' in *Harvard Business Review*, September–October, pp 167–176.

Kearney, R, and E Berman (1999) *Public Sector Performance: Management, Motivation and Measurement*, Westview Press, Oxford.

Keystone and Accountability (2007) *Citizen Sector Accountability: A Review of the Literature, An Overview* by Accountability for the Novartis Foundation (August 2006) Updated by Keystone (April 2007).

Korten, DC (1990) *Getting to the 21st Century: Voluntary Action and the Global Agenda*, West Hartford CT: Kumarian Press.

Lawrie, G, D Kalff and H Andersen (2005) "Balanced Scorecard and Results-Based Management: Convergent Performance Management Systems" Presented at 3rd Annual Conference on Performance Measurement and Management Control, The European Institute for Advanced Studies in Management (EIASM), Nice, France.

Leader, N (2000) *The Politics of Principle: The Principles of Humanitarian Action in Practice*, Humanitarian Policy Group, Overseas Development Institute.

Macpherson, M (1999) *An introduction to ISO 9000 & TQM*, Macpherson Publishing: Alexandra, New Zealand.

Macrae, J (ed.) (2002) *The New Humanitarianisms: A review of trends in global humanitarian action*, HPG Report 11. London: ODI.

Meier, W (2003) *Results-Based Management: Towards A Common Understanding Among Development Cooperation Agencies*, Discussion Paper Prepared for the Canadian International Development Agency, Performance Review Branch, for consideration by the DAC Working Party on Aid Effectiveness and Harmonisation.

Mitchell, J (2008) 'Introduction: Collective action and performance – a personal view' in *ALNAP Seventh Review of Humanitarian Action*. London: Overseas Development Institute.

Mock, N and R Garfield (2007) *Health Tracking for Improved Humanitarian Performance*, Prehospital and Disaster Medicine, September–October 2007.

Murray and Frenk (2006) 'A WHO Framework for Health System Performance Assessment.' WHO Evidence and Information for Policy Paper.

NAO (2001) *Measuring the Performance of Government Departments*, National Audit Office.

NAO (2007) *World Food Programme – Managing for results: A second review of progress in implementing results-based management.*

NGO Coordination Committee in Iraq (NCCI) (2008) *Iraq: Operational Modalities in Iraq.*

Neely, AD, M Kennerley and C Adams (2007) 'Performance Measurement Frameworks: A Review', in *Business Performance Measurement: Unifying Theory and Integrating Practice*, Neely, A.D. (ed), Second Edition, Cambridge University Press, Cambridge.

Neely, AD (2007) *Business Performance Measurement: Unifying Theory and Integrating Practice*, Second Edition, Cambridge University Press, Cambridge.

Neely, A, B Yaghi, N Youell (2008) *Enterprise Performance Management: The Global State of the Art*, Cranfield Centre for Business Performance: Cranfield University.

Obser, A (2007) *Multilateral Organisations Performance Assessment: Opportunities and Limitations for Harmonisation among Development Agencies*, German Development Institute.

OECD (2000) *Results Based Management In The Development Co-Operation Agencies: A Review Of Experience*, draft Report Presented to the February 2000 meeting of the DAC Working Party on Aid Evaluation.

OECD-DAC (1999) *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies*, Evaluation and Aid Effectiveness Series 1. Paris: OECD-DAC.

OECD-DAC (2000) *Glossary of Key Terms in Evaluation and Results Based Management*. Paris: OECD.

OECD-DAC (2005) *Paris Declaration on Aid Effectiveness: Ownership, Harmonisation, Alignment, Results and Mutual Accountability*, High Level Forum, Paris, 28 February–2 March.

OECD-DAC (2008) 'Revised Humanitarian Assessment Framework for DAC Peer Reviews', DCD/DAC, 48.

OIOS (2007) Inspection on Results-based management (RBM) practices at the United Nations Office for the Coordination of Humanitarian Affairs (OCHA).

OIOS (2008) Review of results-based management at the United Nations: Report of the Office of Internal Oversight Services.

Otley, DT (2002) 'Measuring performance: the accounting perspective' in "Business Performance Measurement: Unifying Theory and Integrating Practice", Neely, A.D. (ed), Second Edition, Cambridge University Press, Cambridge.

Oxfam (2009) *The Right to Survive*, Oxford: OXFAM.

OUP (Oxford University Press) (2008) *Concise Oxford English Dictionary* (eleventh edition, revised). Oxford: OUP.

Parmenter, D (2007) *Key Performance Indicators: Developing, Implementing and Using Winning KPIs.* Hoboken NJ: Wiley.

Pollit, C (2001) 'Clarifying Convergence' in *Public Management Review* 4.1, pp. 471–492.

Pollitt, C, and G Bouckaert (2004) *Public Management Reform: A comparative Analysis*, Oxford University Press.

Porter, T (2002) *An External Review of the CAP*, commissioned by OCHA's Evaluation and Studies Unit, April.

Sandison, P (2006) 'The Utilisation of Evaluations' in *ALNAP Review of Humanitarian Action: Evaluation Utilisation*. London: Overseas Development Institute.

Seaman, J (1999) cited in The use of mortality and nutrition indicators: Advantages and disadvantages, WHO Emergency and Humanitarian Action, presentation given in July 2002, downloaded from http://www.smartindicators.org/workshop/pres_smart/July_23_Materials/WHO%207-23.ppt on 25th June 2009 at 0929.

Scott, A (2005) *DFID's Assessment of Multilateral Organisational Effectiveness: An Overview of Results.* London: International Division Advisory Department, DFID.

Smith, A (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, London: Methuen and Co.

Stoddard, A (2003) *Humanitarian NGOs: Challenges and Trends.* HPG Briefing 12, London: ODI.

Stoddard, A, et al. (2007) 'Cluster Approach Evaluation Report Final Draft', 21 November. Geneva: OCHA.

Taylor, FW (1911) *The Principles of Scientific Management.* New York: Harper.

Telford, J, and J Cosgrave (2006) *Joint Evaluation of the International Response to the Indian Ocean Tsunami.* London: Tsunami Evaluation Coalition.

Thiel and Leeuw (2002) '*The Performance Paradox in the Public Sector*'. Public Performance and Management Review. 25(3): 267–281.

UNDESA (2003) Organisational Performance Management and Measurement, Beirut.

UNHCR (2007) 'Health Information System (HIS): A Training Manual to Support Implementation in Refugee Operations' (March). Geneva: UNHCR.

Vaux, T, et al. (2001) *Independent Evaluation: The DEC Response to the Earthquake in Gujarat*. London: DEC.

Wallace, T, with L Bornstein and J Chapman (2006) T*he Aid Chain: Coercion and Commitment in Development NGOs.* London: Practical Action.

Waring, S (1994) *Taylorism Transformed*, UNC Press.

Wood, A, R Apthorpe and J Borton (eds) (2001) *Evaluating International Humanitarian Action: Reflections from Practitioners.* London: Zed Books/ALNAP.

World Bank (2008) *Sourcebook on Emerging Good Practice in Managing for Development Results*, World Bank.