Friday 19 November (14:00–16:00 - HEC amphi 4)

“Compromising accounts: performance measurement, modes of evaluation and the reconciliation of world-views”

Matthew Hall
(London School of Economics)
Compromising accounts: performance measurement, modes of evaluation and the reconciliation of world-views

Robert H. Chenhall
Department of Accounting and Finance
Monash University

Matthew Hall
Department of Accounting
London School of Economics and Political Science

David Smith
Department of Accounting and Finance
Monash University

Preliminary draft: 16 August 2010
Please do not quote or cite this work without permission of the authors.

Acknowledgements: The support of CIMA General Charitable Trust is gratefully acknowledged.

1 Corresponding author: m.r.hall@lse.ac.uk
Compromising accounts: performance measurement, modes of evaluation and the reconciliation of world-views

Abstract

In this paper we outline a distinctive approach to the analysis of accounting and organizational change. Rather than examine how accounting is implicated in the effective replacement of one culture by another (c.f., Dent, 1991; Ezzamel et al., 2008), we focus on the role of accounting in situations where the co-existence of and reconciliation between different world-views is an ongoing feature of organizations and the need for acceptable compromise is paramount. While prior studies typically focus on the characteristics of existing performance measurement systems, we examine explicitly the processes that take place in negotiating the scope, design and operation of performance metrics. In a field study of a non-government organization, we show how struggles over the choice of different metrics, images and words in an account can affect efforts to seek compromise between different modes of evaluation and world-views.
Compromising accounts: performance measurement, modes of evaluation and the reconciliation of world-views

Introduction

“The old world view, the preoccupation with engineering and logistics, the belief in the railway as a social service, the railway culture, has been substantially displaced by the business perspective, the belief that railways should be instrumental in making profit and managed to that end. The counter-culture has emerged to become the dominant-culture among the senior management” (Dent, 1991: 715).

“There's still a big debate in VSO about whether the purpose is to make sure volunteers have a good experience overseas and then return back happy or do we have some sort of coherent development programmes which use volunteers as a key input. I think there are those two different views of the organization. I mean there's a whole lot of views between those two but those are the two extremes... it probably divides down the middle” (Regional Director 2, Voluntary Service Overseas).

The role of accounting practices in changing organizational realities has been a prominent feature in studies of accounting and organizations. Dent (1991), in his pioneering study, showed how accounting techniques and their associated discourses were mobilized by a group of business managers to move a railway company from its ‘railway’ past to its ‘business’ future. Oakes et al. (1998) revealed how accounting and planning techniques were used to alter the conception of an Alberta museum away from its traditional imperative of providing a cultural experience to one focused more on efficiency and the market. More recently, Ezzamel et al. (2008), in their study of a large chemical company, explored the role of accounting calculations in transforming ‘Conglom’ into an organization where operational activities and staff culture were focused on ‘making the numbers’. These studies show how accounting practices were linked with, tightly coupled to and ultimately helped to constitute the emergence of a particular model of organizational reality, one that emerged as the ‘victor’ with its preferred model of rationality and associated accounting practices displacing the old. Although the organization was a site of struggle and resistance, we see how a previously taken-for-granted view of organizational reality receded into obscurity with its preferred model of rationality and associated accounting practices (seemingly) consigned to history. Notwithstanding the importance of these situations, groups within organizations can hold...
differing views of organizational reality that persist over time (that is, remain ‘divided down the middle’) without a shift to a favoured and ultimately dominant world view. It is within this context that we investigate the following questions. What is the role of accounting practices in situations where the co-existence of different world-views is a common, and ongoing, feature of organizational life? How is accounting implicated, not in processes of domination and displacement, but in acts of reconciliation and compromise?

We argue that answers to these questions can advance our understanding of accounting in at least three ways. First, whilst very illuminating, prior studies that link accounting practices to different organizational realities investigate a somewhat limited domain because the object of analysis is how accounting is implicated in the effective replacement of one reality by another. In contrast, the increasing plurality of interests in organizations, including the rise of ‘stakeholder’ models (e.g., Mitchell et al., 1997), indicates that accounting is also likely to be implicated in processes of achieving (or not) co-existence between different ‘worlds’, modes of evaluation and their associated practices. This focuses attention on the ordering of priorities amongst different ‘worlds’ rather than on one world moving to dominate and replace another (Boltanski & Thévenot, 1999, 2006). This perspective privileges a focus on how actors achieve reconciliation and compromise despite potentially divergent and even competing values (Boltanski & Thévenot, 1999, 2006; Thévenot, 2001; Denis et al., 2007; Stark, 2009). Accounts are particularly important in these processes because they are a site where multiple modes of evaluation all potentially operate at once, with different modes of evaluation privileging particular metrics, measuring instruments and proofs of worth (Stark, 2009). Here, accounts can act as a ‘convention’ that crystallizes the compromise between different modes of evaluation and thus enables the reconciliation of potentially competing value frameworks (Denis et al., 2007; Thévenot, 2001).

Second, we argue for a focus on the actual processes of reconciliation that attempt to
establish (or undermine) an inherently temporary co-existence between competing modes of evaluation. Such an approach can improve understanding of how actors with different orders of worth and subsequently differing modes of evaluation reach an acceptable compromise, the factors that promote and/or discredit the compromise, and the subsequent consequences for those individuals, groups, organizations and institutions involved. Variations in the design and operation of accounting practices are likely to be important in whether different worlds are pushed apart or do indeed achieve reconciliation and co-existence. In this way, different modes of evaluation can be viewed as different ways of conceptualizing the representation of performance, that is, through particular types of metrics (e.g., financial vs. non-financial), images (graphs, tables, or photographs) and narrative content (stories, management reports, disclosures). Accounts of performance are critical because it is in discussions over the different metrics, images and words that can be used to represent performance that the actual worth of things is frequently debated, established and contested. An analysis of compromising accounts, and the different modes of evaluation and orders of worth that they bring together, provides a powerful analytical lens for examining whether and how compromise and co-existence between different groups is developed, established and destroyed.

Third, while there has been a wealth of prior management accounting studies focusing on the attributes of various performance metrics and their effects on individual and organizational performance (see for example, research on subjectivity (Gibbs et al., 2004; Moers, 2005), comprehensiveness (Hall, 2008) and financial/non-financial measures (e.g. Perera et al., 1997; Baines & Langfield-Smith, 2003)), most of these studies do not explicitly investigate how the metrics that comprise performance measurement systems are developed, established and contested (see Wouters & Wilderom (2008) and Townley et al. (2003) for exceptions). Thus, we extend this literature by examining explicitly the processes that take

2 We use the term ‘compromising accounts’ to refer to the role of accounts in assisting change by reconciling different world-views. We revisit this concept later in the paper.
place in negotiating the scope, design and operation of the metrics included in performance measurement systems.

The remainder of the paper is structured as follows: In the next section we provide the theoretical background to the study. The third section details the research method, with the fourth section presenting findings from our field study of Voluntary Service Overseas. In the final section we discuss our findings and provide concluding comments.

**Background literature**

Our interest in this study is how accounting is implicated in situations where the co-existence of differing world-views is an on-going feature of organizations. Denis et al. (2007) argue that the processes underlying the attainment of compromise between competing values systems have not been analyzed extensively. We draw on ideas from Boltanski & Thévenot (1999, 2006) and Stark (2009), specifically the concept of ‘orders of worth’, to illustrate how competing world-views are taken into account, or balanced, when parties seek to reach agreement or resolve disputes. A ‘clash’ may emerge between parties, who at a given point in time, and in relation to a given decision, prioritize different world-views.

Boltanski & Thévenot (1999) argue that in seeking to resolve such clashes and gaining agreement the parties need to focus on a principle of equivalence. That is, in order to evaluate the claims made by one party with a particular world-view against the claims of another, it is necessary to have some means of being able to compare, or evaluate, the respective claims. Stark (2009: 12) writes “As coherent principles of evaluation, each of the orders of worth has distinctive and incommensurable principles of equivalence. Each defines the good, the just, the fair – but according to different criteria. Each qualifies persons and objects with a distinctive grammar and logic.” It is here that the processes of accounting, and the potential of performance measurement systems to both capture and emphasize differing modes of
evaluation, become important. Denis et al. (2007: 192) argue that conventions³ “…make the co-existence of heterogeneous worlds possible by providing an acceptable compromise between competing value frameworks.” Because a single individual may identify with multiple worlds, and because different individuals within an organization may not identify with the same worlds to the same extent, the negotiation and development of conventions is seen as a critical tool to aid compromise (Denis et al., 2007). Denis et al. (2007) also argue that conventions can also be a source of tension if people feel that some fundamental principles associated with a world with which they identify are not respected. In this situation, such people may question the legitimacy of the convention (such as performance measurement systems), potentially reducing the likelihood of its success as a tool in aiding compromise.⁴

In the context of an account that seeks to reconcile different world-views, it is the process of deciding the commensurability of the items to be included in the account that is most likely to bring the different modes of evaluation into conflict. Espeland & Stevens (1998: 314) define commensuration as “the transformation of different qualities into a common metric”. It is this transformation that “allows people to quickly grasp, represent, and compare differences...offers standardized ways of constructing proxies for uncertain and elusive qualities...condenses and reduces the amount of information people have to process, which is useful for representing value and simplifying decision-making” (Espeland & Stevens, 1998: 316). Espeland & Sauder (2007) argue that there are three core features of commensuration that influence sense making: commensuration’s capacity to reduce, simplify,
and integrate information, its ability to unite and distinguish between objects\(^5\), and its capacity to invite reflection on what numbers represent.

Actors with different world-views are likely to put forth claims about incommensurables, that is, to “deny that the value of two things is comparable” (Espeland & Stevens, 1998: 326). This is because “incommensurables can be vital expressions of core values, signaling to people how they should act toward those things” (Espeland & Stevens, 1998: 327). Thus, claims over incommensurables are not mere technicalities (although they may be expressed as such) but can embody deeply held values intrinsic to a particular mode of evaluation and order of worth. Arguably, everything is potentially measurable, if we ignore the quality of potential proxy measures. However, objects can be considered incommensurable when the object is so culturally or morally significant that to seek to place a value on it is seen as repugnant, or devaluing of the object under consideration (Raz, 1986; Espeland & Stevens, 1998).

**Method**

We conducted a field study at Voluntary Service Overseas (VSO) during the period July 2008 to August 2010. VSO is a non-governmental international development organization that works by (mainly) linking volunteers with partner organizations in developing countries. Each year approximately 1500 volunteers are recruited and take up placements in one of the over forty countries in which VSO operates programmes, mainly in South and South-east Asia, and sub-Saharan and West Africa. Our interest in VSO as a research site was sparked when we became aware of an initiative to develop a new performance measurement system, subsequently referred to as the ‘Quality Framework’ (QF). This framework attempted to

---

\(^5\) In elaborating on this idea, Espeland & Sauder (2007) argue that commensuration unites and distinguishes between objects insofar as it constructs a precise hierarchical relationship between objects and makes the differences between them less visible. An outcome of this, according to Espeland & Sauder, is that small differences end up being closely scrutinized because the nature of commensuration is such that it makes comparisons of this nature impossible to avoid.
combine different metrics, images and narrative content into a single report that would provide a common measure of performance in each of the country programmes operated by VSO.

As part of the field study, we conducted 32 interviews, attended meetings, observed day-to-day work practices, collected internal and publicly available documents, participated in lunches and after-work drinks with staff and volunteers, primarily in London, but also during a one-week visit to the programme office in Sri Lanka in January 2009.

Most of the interviews were conducted by one of the authors, with two of the authors conducting the interviews with the country directors. Interviews lasted from 30 minutes to two hours. Almost all interviews were digitally recorded and transcribed, and, where this was not possible, extensive notes were taken during the interview and further notes then written-up on the same day. We interviewed staff across many levels of the organization as well as staff at different locations. Face-to-face interviews were conducted at VSO headquarters in London, and during the one-week visit to Sri Lanka in January 2009. Due to the location of VSO staff around the world, some of the interviews (particularly with country directors) were conducted via telephone. An overview of the formal interviews and observations of meetings is provided in Table 1.

We carried out observations of 17 meetings and workshops in both London and Sri Lanka. In London we attended meetings related to the QF, as well as other planning and review meetings. In Sri Lanka, during the five day visit in January 2009, one of the authors observed office planning and logistics meetings, as well as a one-and-a-half day planning and review workshop.

Throughout the study, we were also involved in informal conversations, usually before and after meetings, as well as during coffee breaks, lunches and after-work drinks. During these conversations staff and volunteers expressed their thoughts about the meetings, as well
as other goings-on at VSO and in the non-government organization sector more broadly. We kept a detailed notebook of these informal observations, which included verbatim statements as well as notes on interactions between staff, ad hoc comments and opinions, and the timing, location and layout of meeting rooms. As soon as possible after each visit, the notes were written up into an ‘expanded account’ (Spradley, 1980) that on completion of the field study totalled more than 200 pages of text. We also exchanged numerous emails (over 700 separate communications) and telephone conversations with staff at VSO.

We were provided access to over 600 internal VSO documents, including performance measurement reports with supporting documents and analysis. These reports included the complete set of ‘QF’ reports from each VSO programme office for the years 2008 and 2009, documents related to other monitoring and review processes, as well as more general documents concerning organizational policies, plans and strategies. Finally, we collected publicly available documents, such as annual reports and programme reviews, case studies, press releases, newspaper articles, as well as several books on the history of VSO.6

Consistent with the approach employed by Ahrens & Chapman (2004) and Free (2008), we employed Eisenhardt’s (1989) methods. This involved arranging the data (transcripts, field notes, documents) chronologically and identifying common themes and emerging patterns. We then re-organized this original data around key events and significant issues as they related to our understanding of performance measurement and review systems at VSO. We compared our emerging findings from the study with existing research to identify the extent of matching between our data and expectations based on prior theory. In particular, findings that did not appear to fit emerging patterns and/or existing research were highlighted for further investigation as the research progressed. This process was iterative throughout the research, and finished when we believed we had generated a plausible fit between our research questions,

---

6 For example, see Bird (1998).
theory and data (Ahrens & Chapman, 2006).

In presenting our findings, we commence by showing how performance measurement systems based on narrative were developed to assist learning, and, in addition, formal quantitative measures were then adopted to assess effectiveness in delivering development outcomes. We show how different views on VSO’s purpose created tension and how debates on how to measure the value of VSO’s work were implicated in attempts to achieve reconciliation and compromise between these world-views. In this process of change we identify how a performance measurement system helped to develop compromise by reconciling an aim to standardize assessment while respecting uniqueness of operations and a need to inspire at the local level. We show how this compromise was short lived where changes in the use of narrative and quantitative measures produced uncompromising accounts that favoured a world-view based on a need for ‘consistency’ in evaluation over the need for ‘learning’. A final stage in the development of the performance measurement systems involved further changes in order to develop a compromising account that helped to ensure coexistence between competition and learning world-views.

Co-existing worlds in tension

VSO was founded in 1958 in England as an organization to send school leavers, predominately 18 year old males, to teach English in the “underdeveloped territories of the Commonwealth” (Bird, 1998). Volunteers were initially recruited exclusively from England, and then from other countries in the ‘North’, i.e., the Netherlands and Canada. In 2000 VSO opened its first volunteer recruitment base in the ‘South’, with the establishment of VSO
Jitolee in Kenya, followed by further bases in the Philippines and India. Geographic expansion was also accompanied by changes to the age, experience and work of volunteers. The 18 year old high school graduate had been replaced by a (typically) 30 year old-plus experienced professional. Volunteers were unlikely to engage in service delivery during their placements, such as teaching in the classroom, but operated under a capacity building approach, being involved in teacher training, curriculum development, advocacy, as well as building administrative and organizational capacities in Government departments of education. VSO was founded on the basis that volunteering was uniquely connected not to the giving of ‘aid’ but to mutual exchange, an ideal that continued in VSO’s slogan “Sharing skills, changing lives”.

The focus on volunteering and civic engagement had privileged a particular mode of evaluation within VSO; one focused specifically on “learning”. Discourses within VSO regularly emphasized the importance of learning, with one of VSO’s three stated values a “commitment to learning” whereby VSO seeks to “continue to develop effective monitoring and evaluation methods so that we can learn from our own and others’ works” (VSO, 2004). A dedicated team was called “Programme Learning and Advocacy” (PLA), whose main role was to support programme offices in learning from their own work as well as sharing practices between countries. Within programme offices, the annual review process for each programme emphasized learning, where Programme Area Review reports contained dedicated sections to describe “key learning”, as well as providing case studies of specific events or activities, often referred to as “Most Significant Change” stories. Another important feature was the opportunity to share good practice with other programme offices. For example, a

7 VSO operates what it calls a ‘capacity building’ approach by partnering volunteers with local organizations that require assistance or expertise in a variety of capacities. VSO describes its partnership approach as “we work with local partners in the communities we work with, placing volunteers with them to help increase their impact and effectiveness” (VSO website, http://www.vsointernational.org/vso-today/how-we-do-it/, accessed 7 April, 2010). Volunteers typically take up a specific role or position, often working alongside a local staff member, where partner organizations range in size from very small, local businesses, community groups and NGOs, to large organizations and government departments and ministries.
Programme Area Review report from a mental health programme in Sri Lanka suggested that:

“Developing resource partnerships with specific volunteer supply organizations and linking up with other organization to provided international exchanges etc. are useful ways of working more creatively. This way of working could be replicated for other programmes across VSO more widely.”

In fact, the focus on learning was for some staff an important basis for their attachment to VSO:

“That's why I'm still here because I think it's a learning organization…there are still a lot of things we need to improve on in terms of monitoring and evaluation but the fact that VSO is serious in pursuing it that makes me think that it has always been a learning organization or at least the time that I've been with VSO and I really appreciate that.” (Interview, Country Director 3).

In 2004 VSO signalled that it would adopt a more ‘programmatic’ approach to its work, with the release of its new strategic plan “Focus for Change”. This approach meant a movement away from the model of work that was centred around each volunteer placement, to one that meant “focusing all our efforts on achieving specific development priorities within the framework of six development goals. These goals reflect internationally-agreed development targets, our partner organizations’ priorities and areas where we know that VSO can be most effective” (VSO, 2004). The movement to a predominantly capacity building and programmatic model was coupled with explicit recognition of VSO’s purpose as primarily a ‘development’ rather than ‘volunteer-sending’ organization, and the development of systems to support this change:

“We are currently developing monitoring and evaluation tools so we can assess changes in the capacity of our partners, establish whether disadvantaged people consider their situation has improved, and identify VSO’s contribution to that improvement” (VSO, 2004).

Notwithstanding this explicit shift in organizational priorities, the focus on volunteering was still strong, particularly as many VSO staff were former volunteers. As such, a mix of different world-views at VSO was the norm:

“There are some different kind of ideological views between people who feel that the

---

9 The six development goals in Focus for Change were health, education, secure livelihoods, HIV/AIDS, disability, and participation and governance.
important thing about VSO, it's just about international cooperation and getting people from different countries mixing with each other and sharing ideas. It doesn't matter what the outcome is really, it's going to be a positive thing but you don't need to pin it down. Versus it's all about pinning down the impact and the outcomes of our work and being very focused and targeted and being able to work out what is your return on your investment and all these kind of things so I think it is partly historical and partly differences in just a mindset or world view.” (Interview, Regional Director 2).

These different views on the overall purpose of VSO created considerable tension, focused in particular on debates about the value of VSO’s work. Originating from its founding principles, many staff and volunteers felt that volunteering was, in and of itself, a positive and productive activity and any drive to specify an ‘outcome’ of this was secondary. In contrast, the move towards a more programmatic approach, coupled with the recruitment of many staff from other international development agencies, gave much more attention to poverty reduction and demonstration of the ‘impact’ of VSO’s work in developing countries. It is in the context of this tension that attempts at reconciliation and compromise between these differing world-views were made, and was most acute in debates about how to measure the value of VSO’s work.

The move to be more programmatic placed increased pressure on the allocation of resources, both money and central support services, amongst programme offices. Each of the approximately 40 programme offices was a budget holder, and received core funding from VSO headquarters via the annual budgeting process. Concerns over resource allocation were most acute in discussions about the opening and closing of programme offices:

“Focus for Change has built into it a logic which is going to make it more expensive to run our country programmes. We're asking for quality which is going to result in the overhead infrastructure around a programme becoming greater. You're not going to have a Country Director running a programme of 50 volunteers single handedly. That's built into it that you're going to have to take on more staff to get the quality you're looking for. To do that,

10 VSO operated a geographic structure, whereby several programme offices were grouped together to form a specific region, for example, Sri Lanka, India, Bangladesh, Nepal and Pakistan formed the ‘South Asia’ region. Each Country Director reported to a 'Regional Director', with the Regional Directors reporting to the Director of IPG, based in London. IPG in London also had staff responsible for providing support to programme offices in areas such as funding, advocacy, and programme learning and reporting. An overview of VSO’s structure is provided in Figure 1.

11 Core funding related to costs such as staff salaries and benefits, office and vehicle rental, and volunteer costs (including allowances and training/support costs). Each programme office received a limited amount of funding for ‘programme’ costs, with programme offices expected to apply for grants from donors to support further programme work.
we're going to have to make difficult decisions at a global level about which countries we think are really important.” (Interview, Senior Manager, IPG).

However, making these decisions had indeed proved difficult in the past, with the number of programme offices changing each year, with some programme offices opening one year only to close several years later. The lack of a mechanism for evaluating the effectiveness of each programme office was viewed as the key problem limiting a consistent and more rigorous analysis. In particular, the view was that performance was typically based not on evidence but on anecdote or self-promotion:

“One of my observations was that people would say things like “Oh X country is really in trouble or E country is fantastic or C country…” and it concerned me. I used to say, “But what are you basing this on? What's your rationale and if a country does have a problem what is it that is a problem and how could you address it?” Similarly with success because it seemed to me that if you are very noisy about sharing your stuff, which some countries are very good at, then you got a reputation for being good at everything and you'd always be the star country but it wasn't really based on a great deal of fact.” (Interview, Country Director 4).

To move away from this situation, over time VSO had instigated three approaches to measuring the performance of its programme offices, which we detail below. The first approach focused on quantitative scoring and ranking. The second approach focused on narrative and stories. Both approaches privileged a particular world view to the exclusion of others. The third approach (and our main empirical focus) attempted to compromise between these two approaches and thus produce an account focused on the co-existence of, and reconciliation between, different worlds.

<insert Figure 1 here>

Uncompromising accounts

In 2002 VSO developed the “Strategic Resource Allocation” (SRA) tool as its first attempt to measure the effectiveness of each programme office.12 In contrast to the predominant use of stories, case studies and images in much of VSO’s reporting processes, the SRA was based almost exclusively on using numbers to represent performance. For

---
12 See Appendix 2 which provides the ‘summary page’ of the SRA.
example, each programme office was required to score itself on 16 criteria related to the extent to which its work was focused on disadvantage, achieved certain outputs related to volunteers, and adopted a strategic approach. Each criterion was given a precise percentage weighting, e.g., 2% or 4% or 17%. Scores on the 16 criteria were to be aggregated with each programme office awarded a percentage score out of 100, with clear recognition that “the higher the overall percentage a Programme Office receives in this tool, the more “effective” it will be perceived to be based on this measure.”

The scoring process further pushed the SRA away from ‘learning’ towards ideals of consistency and comparability. Programme offices were provided with a guidance document that contained detailed instructions on the process, evidence and even types of judgements to be used in scoring each criterion. For example, scoring the second criterion entailed a four-step process, where step two listed nine key questions that formed the basis for gathering evidence to decide and justify the resultant score, and step three required a conversion of such evidence into a score out of 10, with descriptions provided as to what performance at each of the 10 levels looked like. In addition, there was a strong emphasis on review of scores by staff in London “to ensure consistency between regions, and to make any amendments necessary in light of these comparisons…[to ensure London staff] have a clear idea of the rationale behind the scoring, in order to ensure transparency and to allow comparison between countries.”

The desire to find a more robust and rational way to compare the performance of programme offices was to be achieved through a focus on consistency and standardization. However, a privileging of these ideals was not consistent with the mode of evaluation in use in most of VSO’s review processes, i.e., those focused on learning.

“The SRA was dropped because it was becoming increasingly apparent that some programmes were rather self-critical while others were not - but that this did not necessarily relate very closely to programme quality - in fact it appeared that sometimes the opposite was true, the programmes that had the capacity to critically assess their own performance (and

The self-critical capacity of many country directors developed from the emphasis at VSO on reflection, being explicit about difficulties, and learning from mistakes. In effect, being self-critical and giving oneself ‘low’ scores was considered an appropriate practice in a mode of evaluation where learning was the primary aim. In contrast, in a mode of evaluation where consistency and comparability are crucial, this self-critical capacity could not only lead to programme closure, but its prevalence amongst country directors and regional directors ultimately led to the SRA’s demise. Here, we see that the design and operation of the SRA was focused exclusively on quantitative scoring in a consistent and standardized way. This approach was completely at odds with existing evaluation practices based on learning. In effect, the SRA privileged a single mode of evaluation such that compromise between the differing world-views was not possible.

However, the underlying need for comparisons of programme offices remained. As such, subsequent to the SRA, a new reporting format was developed - the “Annual Country Report” (ACR). Through a narrative report, the ACR was focused on reporting on progress towards ‘Strategic objectives’, contained one or more ‘Most Significant Change’ stories, and also a section called ‘Lessons’. Such an approach fitted squarely with VSO’s focus on learning and reflection in other review processes. In contrast to the SRA, the ACR contained no quantitative scoring or ranking of programme offices. The lack of readily comparable data in the ACR meant that there was still a lack of evidence on which to build a rigorous comparison of programme offices. Commenting on the situation that existed during the use of the ACR, a country director explained:

“I mean VSO doesn't have any objective criteria on the opening and closing of programmes. If you ask anybody at VSO why do we open or why do we close programmes I mean there's no intelligent answer to that. Because there has been no way to make a comparative estimate

of value because there has not been a good tool. I mean if you're going to make decisions around opening or closing programmes there has to be a comparative analysis.” (Interview, Country Director 2).

Here, we see that the design and operation of the ACR was focused exclusively on learning and reflection and its lack of any standardized measures of performance precluded any ‘comparative analysis’ of the effectiveness of programme offices. In effect, the ACR also privileged a single mode of evaluation (that of learning) such that compromise between the differing world-views was still not forthcoming.

**Developing a compromising account**

In early 2007 a new Director of the International Programmes Group (IPG) was appointed. He immediately started a process to develop a new tool for measuring the performance of programme offices, which came to be called the “Quality Framework” (QF). The imperative for compromise between VSO’s different modes of evaluation in the development of the QF was revealed in the first section of an email sent to all country directors by the Director of the IPG in May 2007:

> “The purpose of this message is to propose a process for ensuring that between us all we can do some work defining what a successful programme looks like for VSO - and then determine ways of actually measuring that success...I believe that it is absolutely essential that we have a shared vision of success - that we all know what a high quality, successful VSO country programme could look like - that we know how to measure this - and that we have a culture that encourages, supports and celebrates this. Of course all of our country programmes could, and should, look very different. Local circumstances and development priorities above all should ensure this...However, there must be some fundamental principles that drive VSO's programme work and enable us to determine whether we are successful or not.”

The reference to a ‘shared vision’ and ‘fundamental principles’ implies common and standardized ways of measuring success, but at the same time recognizing the uniqueness of country programmes in that they ‘should look very different’. The task, then, was to move beyond the SRA and ACR and to design an account that could enact a compromise between the different world-views and the modes of evaluation that underpin them.
How to standardize yet inspire?

A key difficulty in developing the QF was trying to reconcile between the need to standardize, i.e., have indicators that provide a consistent method for measuring success in each programme office, whilst concurrently respecting the uniqueness of programme offices and the need for indicators to be ‘inspirational’. It is at the specific design level, i.e., making choices about the content of elements, indicators and narrative components in the QF, that the compromise between different modes of evaluation was most acute. Importantly, country directors were viewed as central to this task and given the opportunity to make suggestions for elements and indicators. The IPG Director explained the process in an email to country directors in May 2007:

“It is being proposed that we seek to develop a Quality Framework with key indicators for VSO’s programmes. Such a framework would enable us to determine the following:

- Indicators of success in operational practice
- Good practice in ways of working
- Quality in programme outcomes

In effect we would develop a simple list of the key indicators that we would all look at to help determine whether a programme is successful…this process will involve others across the organisation…For now, however, we want CDs (with input from your programmes) to start determining the key indicators.”

Input from country directors and programme staff was collected in London, and then followed in late 2007 by a meeting of all country directors and senior IPG staff in Cambridge, UK. A central platform of this meeting was sessions devoted to dialogue and debate about the elements and indicators that would comprise the QF. Centred around the question “What is quality?”, it was here that staff with different world-views were able to advocate for the inclusion and exclusion of particular elements and indicators. This resulted in a set of 14 elements relating to various aspects of programme quality, such as inclusion, volunteer engagement, innovative programming and financial management (see Appendix 2). Each element had a set of indicators (from a low of one indicator to a high of five indicators per element), with the scores on the indicators, along with country director and regional director
judgement, used to determine an overall element score.

Local knowledge was considered paramount in that indicators were required to produce results that were ‘recognizable’ to programme office staff, partners and volunteers. Providing space in the QF report for narrative discussion to reflect the different operating environments in different countries was also critical. Reflecting on the ‘pilot test’ of the QF, a country director commented:

“We worked through the different indicators to see whether the results that the framework spat out were recognizable…we went through and did it and then I looked at whether I recognized the result that it gave then [IPG staff member] also spoke to staff and different partners and some volunteers to see whether they recognized those results and some of the results didn't give the right answer basically. So we changed some of the indicators...The framework itself allows for a small narrative at the beginning of each element. Which can at least explain context as to why it may have a low score or conversely why it might have a high score. They may be working in a country that has a very easy operational environment. It might have lots of external funding and that for me is reflected in that short narrative section at the beginning.” (Interview, Country Director 1).

The QF also ensured that there was flexibility in the design of the performance ranges for the indicators:

“Many of the KPIs have got ranges set against them to outline what ‘high performance’, ‘satisfactory performance’ and ‘room for improvement’ looks like. However, in some cases it will be more relevant for CDs and Regional Directors to decide what results say about the performance of the programme within the context of the programme itself…it is recognised that what can be considered high performance will differ considerably between Programme Offices.”\(^\text{16}\)

As such, eight of the indicators were scored with guidance explaining that “performance [to be] determined by PO [programme office] and RD [regional director]” rather than the use of explicit performance ranges. Unlike the SRA, the QF did not include elements related to advocacy and learning or give elements explicit weights. This meant that the production of an overall score for a programme office was not a feature of the QF (see Appendices 1 and 2 for a comparison of the elements in the SRA and QF).

Collectively, the input of country directors and other programme staff, the privileging of local knowledge in designing indicators, providing space for narrative so that programme offices could reflect local circumstances, and recognition that performance on some indicators

was best determined using programme office and regional director judgement, provided explicit recognition of the uniqueness of programme offices, and the message that the QF was not just about scoring. Additionally, discussion around the elements of the QF, particularly those related to what was to be included and excluded, represented a rich opportunity for constructive debate over areas of prioritization. Central to this debate was discussion around the ‘measurability’ of certain aspects of performance, with some staff arguing that certain aspects of performance (for example, indicators for the “Inclusion” element) should not be scored, as their very nature did not lend itself to quantification. To some extent, concern over such issues was alleviated by the ability to ‘include’ such items in the space devoted to narrative, which country directors felt was a critical component of the QF as it allowed discussion of important issues that they believed were not captured in the indicators. For example, one country director commented that “I don't think the scoring was very useful…In my narrative I was able to clearly say these are the big issues for us”. Critically, however, the need for comparability was also recognized in the QF. Each programme office was required to complete the QF using a common template, with common elements and indicators, thus providing a standardized method of measuring performance across countries.

After its first year of operation, praise for the QF was widespread. Many country directors commented that it helped to provide clarity around what made a quality country programme and was also viewed as better than the ACR:

“Overall it was a good move away from the Annual Country Report because one of the main things was it gave much more direction on being clear on what to report on but also through the report it identified what is important, what’s quality but it's also important to reflect on as a programme. Now you can always argue about elements of that, that's not the point. I think it's just helpful to say well these are overall important parts to reflect on and I thought that was quite useful.” (Interview, Country Director 2).

There was also scope for constructive debate lasting beyond the initial development of the QF through the requirement for country directors and regional directors to both complete scoring for an individual programme office, and to discuss differences in scores that emerged
from this process. Many staff from programme offices completed the QF together, providing a way to review overall results, and many regional directors used the QF to help set objectives and action plans for country directors in the coming year. Some programme offices embraced the QF even further, using it to determine whether an office move would increase programme quality or further disaggregating the QF so it could be applied to different parts of the programme office. The Director of IPG also praised the QF, particularly the way it had shifted the focus of staff in programme offices away from volunteer numbers or funding levels to the impact of programmes on partners and beneficiaries. At the same time, the ability to aggregate data across programme offices was viewed as critical in being able to provide evidence in arguments for more resources:

“I think it's been great. It's not a perfect tool but I don't think any tool in development ever is perfect… there wasn't a lot of discussion about quality or about success and the discussions were more about how many volunteers have we got or how much programme funding have we got and the quality framework has been a really useful tool over the last 18 months for just getting people to talk more about impacts on poverty. Quality, what is quality like?…To just get that narrative going across programmes, across the organization so that people are just using that kind of language and thinking in that kind of way…[The QF has] given me stronger evidence when arguing at senior management team level for where things aren't working. So when you've got 35 country directors saying things like the finance systems aren’t working it gives you a lot of evidence to be able to really argue for that… for more resources to go into strengthening the finance systems… so from all of that basis, I think it's gone really well.” (Interview, Director, IPG).

Statements that the QF is not ‘a perfect tool’ but is ‘quite useful’ and ‘worked out pretty well’ reveal an awareness of the compromises in play, and helped to enact a certain stability between a mode of evaluation that privileges local context and uniqueness and one that favours standardization and comparability. But such stability was temporary. Shortly after the completion of the QF in its first year, the compromise was subject to strong criticism, aimed in particular at the processes used to score the indicators and elements.

---

17 The Director of IPG was a member of the six-person executive management team at VSO called the ‘Senior Management Team’ (other members were the Chief Executive Officer, Director of the VSO Federation, Director of the UK Federation, the Chief Financial Officer, and the Director of Human Resources). This group was responsible for major resource allocation decisions, particularly the amount of funds that were allocated to each of the major divisions within VSO, including IPG. As such, the total amount of funds allocated to IPG was due, at least in part, to the ability of the IPG Director to put forward a solid case for more money as this would inevitably mean that other divisions would receive less money.
But you’re not consistent!

The scoring process for the QF was based on self-assessment by the country director (with programme office staff input), with review by the relevant regional director. This raised concerns, particularly from some staff in London, that scoring was not consistent across regions and countries:

“Key anomaly is that the ratings seem to have been applied differently in each region…I believe this is an inaccurate reflection of the current strengths and weaknesses of programme funding across IPG…I suspect there are different interpretations as to what constitutes good programme funding performance…I think there is a need to clarify what justifies a 1, 2, 3 or 4 within each indicator.”\(^{18}\)

A senior IPG manager was the most vocal critic of inconsistency in the scoring process:

“A 3 in one place means something that could be completely different from a 3 on the same criterion in another country…I think both the indicators against those elements and the process for arriving at a score are dubious. I would even go so far as to say extremely dubious.”

He viewed taking the scoring of indicators and elements out of the hands of country directors and regional directors as the best way to get consistency. First, he floated the idea of an external assessment process:

“I was saying to the managers the other day, really more for the sake of being provocative, why not let’s just have a sort of OFSTED\(^{19}\) type inspection unit? You don’t have schools self scoring to get their school inspection done. A team come in who are completely external and apply consistent methodology right across the country…Perhaps we should do that.”

then an internal assessment unit:

“There’s an argument for having the whole thing handed over to an independent performance assessment unit within VSO…I was trying to get going a debate about how we can build in more independence. A triangulation of the assessments…Possibly triangulate it with sampling. Get the internal auditor to do three [of the QF reports]. Just as a norming exercise across regions.”

The emphasis on having a ‘consistent methodology’ with ‘independence’, along with a focus on ‘norming’, privileges comparability over local context and country uniqueness.

\(^{18}\) QF Element Summary document, 2008.

\(^{19}\) OFSTED is the Office for Standards in Education, Children’s Services and Skills in the UK, an independent body that inspects schools and provides ratings of performance. For example, schools are awarded an overall grade from 1 to 4, where 1 is outstanding, 2 is good, 3 is satisfactory and 4 is inadequate (OFSTED website, [www.ofsted.gov.uk](http://www.ofsted.gov.uk), accessed 24 July 2010).
Having been employed at VSO during the development of the SRA, he lamented its demise and concluded that the balance in favour of self-assessment in the QF had created questionable results:

“In many ways the scoring methodology [in the SRA] was a hell of a lot more robust than what we've done this time around [with the QF]…arriving at the scores and then having to moderate it on every criteria was a lot subtler and more managed, much more managed…the downside of that was it took a lot of time whereas this exercise [the QF] was pretty light on workload… How much time is it worth putting into this thing? Would it be worth spending two months on it? Absolutely not. You've got to get the balance right. I think you've got the balance too far towards the end of quick and dirty.”

Although the use of either an internal or external performance assessment unit did not materialize, several changes were made to the QF in readiness for its second year of operation at the instigation of staff in London with little consultation with country directors. Each indicator now included a description of each of the 1-4 levels, where previously only levels 1 and 4 had a description. Revised guidance documentation also revealed a renewed emphasis on the importance of scoring:

“It is important to score yourself precisely against the descriptors. There may be very good reasons why you achieve a low score on a particular indicator, but it is important to score precisely – the narrative can be used to give a brief explanation.”

Furthermore, explicit recognition that “what can be considered high performance will differ considerably between Programme Offices” had been removed from the guidance documentation altogether, and the scoring of only one (rather than the previous eight) of the indicators would be assessed using programme office and regional director judgement. This movement away from appreciation of local conditions to standardization was particularly apparent with indicator 8.1, with an ability to assess performance on a ‘country by country’ basis being replaced with explicit monetary ranges that would apply to each programme office regardless of its size of operations or different external funding environments (see Table 2). Ownership of some scoring was also taken away from programme offices:

“Note that for indicators 3.1, 3.2 and 3.3 the scores will be pre-populated in London by

---

extracting numerical data from Starfish\textsuperscript{21}. This allows data to be comparable across programmes by using universal interpretations of the data.\textsuperscript{22}

Concerns over the scoring process itself were also addressed, particularly in relation to ensuring consistency in the way that regional directors used indicator scores to determine overall element scores. In a meeting of regional directors, one commented:

“The bit we haven’t specified very objectively, is how you score the overall elements as an RD. So you get lots of stuff in at the end of April, all these scores are there, what are you going to do with them? How are we going to standardize our scoring a bit? We can think about a way to do that, whether it is pairing people up or some sort of calibration of scoring to make sure we get that consistency” (Regional Director 2, QF meeting).

A further meeting was arranged to address this issue of consistency. Lasting for over two hours, regional directors used a recently completed QF report as an example to work through how to score each element. One by one, through each of the fourteen elements, the process for using indicator scores to determine an overall element score was discussed. Looking fed-up, a regional director countered:

“Can I just ask, do we really care how accurately we score? [quizzical looks from other regional directors]. No, honestly, so we could spend a lot of time working out how we score it and use it for comparison but I mean you could roughly get a score on an average without spending too much time on the scoring but concentrate on what they’re saying, and concentrate on quality discussion which presumably we also want to do.” (Regional Director 5, QF meeting).

The ensuing discussion focused not on what programme offices were ‘saying’ or how to ‘concentrate on quality discussion’, but on whether the ‘average’ was a legitimate way to determine element scores, or whether the most important indicator in each element should be designated as the ‘lead’ indicator for the purposes of scoring. Here, disagreements were focused exclusively on the design of the QF itself, rather than it serving as a mechanism to debate the different world-views in play. Despite a lengthy discussion, at the end of the meeting there was no agreed process, except for general agreement on how not to score the elements:

“At the end of each section of the report, there is room for the RDs to give an assessment of

\textsuperscript{21} Starfish was the system used at VSO for tracking the number of volunteer placements, arrivals and departures from each programme office.

\textsuperscript{22} QF Reporting Template, 2009.
The overall performance against the elements won’t be based on an arithmetic average of KPI results for that element, but will instead reflect the RD’s judgement based on the KPIs and on further discussions with the CD.” (emphasis in original).  

Notwithstanding the recognition of the importance of judgement as the final arbiter and not the ‘average’, virtually all changes to the QF design and scoring process privileged consistency in scoring over country uniqueness. Whilst more consistent scores were likely, the initial compromise appeared tenuous, particularly as those whose mode of evaluation was centered on uniqueness and learning expressed concern that fundamental principles associated with their world were not being respected.

But you’re not inspiring!

The focus on consistency and precision in scoring meant that the use of indicators to inspire was given minimal attention. An internal review of the QF after the second year stated: “there was a feeling that they [the indicators] tend to focus on ‘business as usual’ rather than being ‘stretch’ objectives, which push programme offices and breed innovation and ambition.” Similar concerns were expressed in a meeting of regional directors, where the QF was viewed as a ‘set of indicators’ rather than a tool for emphasizing ambition and creativity:

Regional Director 3: The messages that we’re giving to programs at the moment about thinking creatively, being ambitious, being innovative and so on, are not necessarily captured in this element, in this thing here [points at the QF document]…I think this is really good for telling us where we’re at and measuring what we’re looking at measuring but in terms of really looking to shift and change, I just wonder how we’re going to do that, and where that’s captured.
PLA Staff 1: I think there’s a bit that’s still missing in the quality framework because it’s become a set of indicators, so the bit that I think is missing is that we don’t really have anything about culture.
Regional Director 3: Yeah, that’s what it is, yeah [enthusiastically].
PLA Staff 1: If people fulfill all these indicators…that might not be enough to achieve what we’re really looking for, you know…we’ve got fixed on the elements but there’s something behind it all that we haven’t quite nailed…
(QF meeting).

---

A concern with a lack of inspiration was also coupled with critical feedback on other changes made to the QF. In particular, the emphasis on standardization and lack of appreciation of the unique circumstances of countries was of particular concern for programme offices, with many commenting that this meant that indicators do not capture their performance ‘accurately’. Changes also reflected more general concerns over the way in which uniqueness of the programme office had started to fall victim to the push for ‘consistency’:

“While a certain amount of consistency across countries is important, it should not be at the expense of programme offices being able to react appropriately and effectively to the needs of the country and programme. It sometimes appears that a small number of programme offices are having a problem or not meeting basic standards and then a solution is imposed on all programme offices – even though the majority of offices have their own systems and procedure which are working perfectly well...A more nuanced approach with a focus on learning from programme offices that are doing something well rather than focusing on solving the problem would be more effective”\(^\text{24}\)

Reflecting more general concern with the changes, another programme office commented in their QF report:

“The QF report has grown exponentially this year, and the indicators have changed and were only issued a month before the report is due…Good practice in monitoring and evaluation is to collect evidence and learning on a day-to-day basis, which is difficult to do if the ground keeps shifting under one’s feet.”\(^\text{25}\)

Here, ‘good practice’ in monitoring and evaluation is implicitly connected to ‘learning on a day-to-day basis’; changes made to the QF, however, were directed at ensuring robust, consistent and precise scoring in an annual exercise. Increasingly, the co-existence of and reconciliation between these different modes of evaluation was becoming difficult. A regional director summarized the state-of-play after completion of the QF for second year:

“people can see that we’ve tried to make it a little bit more objective in the way that it’s done, but I am getting quite a lot of critical feedback that the quality framework is so big, so many indicators, stuff being sent really late…the whole thing is just a quantitative scoring tool and it’s not about learning in any way shape or form…so I am getting quite critical feedback” (Regional Director 5, QF meeting).

Thus, we see that an initial compromise between standardization through scoring and recognition of country uniqueness had faltered. In this setting, the initial praise for the QF had dissipated and was replaced by critical feedback, particularly from country directors who felt that the push for consistency had moved too far such that the QF was no longer about learning and inspiration but a ‘quantitative scoring tool’. This was particularly evident in debates about particular indicators, standardized performance ranges, or consistent scoring methodologies, where increasingly we see that discussion focused exclusively on the QF itself, with the previously fruitful discussions about how to improve quality or how to reconcile competing values at a minimum. This move toward a more uncompromising account was also evident in initial debates about how to use the QF to improve programme office performance.

How to learn yet compete?

Given the considerable effort that had gone into the development of the QF, there were high hopes that it would help to improve the performance of programme offices. On the one hand, there was a strong desire to improve through ‘learning’, whereby the QF would help identify examples of good practice that could then be shared amongst programme offices. Concurrently, there was a belief that the QF could be used to generate a sense of ‘competition’ amongst programme offices, which would lead to increased motivation and thus improved performance. These different positions on the avenues to improved performance presented many obstacles to enacting an acceptable compromise; obstacles that, at first, proved difficult to overcome.

Although not stated explicitly in any QF documentation, the engendering of a sense of competition emerged through several ways in which the QF data were used. In particular, in its first year of operation, all the QF results were arrayed in a spreadsheet and distributed to country directors via email:

Country Director 9: We were sent back a kind of world-wide kind of scoring sheet and
obviously all that that had was a series of numbers. Sort of 1, 2, 3 filled with red and green and orange. Although to be honest we came second in the world… I feel quite sorry for the countries that have scored quite low because I really don’t think it’s a valid scoring system…but for me it was quite handy to be able to say this and say “maybe next year- first” and all the rest.

_How do you know you are second in the world? Was there some kind of ranking?_
Country Director 9: Yeah, they sent us a ranking. They sent us a worldwide ranking thing afterwards.

_And so countries were rank ordered from 1 to 34?_
Country Director 9: Yeah it was a summary of results with ranking. And you could look against different scores so you could see that…globally you came second on something and third on something else but then there was an overall sort of score.

(Interview, Country Director 9)

Despite reference to an ‘overall sort of score’, the spreadsheet itself did not contain a summary score for each country. Countries were not ranked, but listed in alphabetical order. As such, Country Director 9’s country only appeared ‘second in the world’ by virtue of its name beginning with a letter near the beginning of the alphabet. Other similar stories emerged of countries with names that started with letters towards the end of the alphabet believing that they had performed poorly on the QF. These examples were the source of much joking at meetings of IPG staff in London, with suggestions that ‘Albania’ will be top of the league table next year, and that the solution was to put countries in reverse alphabetical order.26 However, confusion over the creation of an overall score was not limited to country directors. At a team meeting, a new regional director assumed that countries were given an overall score, a practice emphatically denied by the rest of the team:

Regional Director 1: When you said we have a score for the program office I assumed that was some kind of composite aggregated score, that this country scores two and this country scores three, ‘cause otherwise, what do you do with the end result, what did you do last year with the end results, how do you use them?
Director, IPG: Produce a league table and sack people! [ironic]
Regional Director 1: Well, it’s such a complicated exercise.
Regional Director 2: You’ve got twelve elements, so the key bit you’re focusing on is giving an overall element score.
Regional Director 1: Sorry, underneath my question was that if you rank all of those twelve elements equally and added them all up, to my mind that wasn’t fair because some things are more important than others.
Regional Director 2: We don’t do that.
Regional Director 1: Ok so you don’t do that, so what I’m not clear about is how you get from the twelve individual numbers to some overall grade.
Director, IPG: We don’t.
Regional Director 2: We don’t have an overall grade. It gives you a holistic picture of how the

26 The use of Albania as an example was ironic because even though VSO did not operate there, it would automatically come first in a league table because no other countries that VSO worked in began with ‘A’.
But the light-heartedness about rankings and insistence that the QF was not about giving an ‘overall grade’ belied an appreciation of how aware country directors were of the competitive mantra that lay behind the distribution of the spreadsheet, and how this was viewed as stifling opportunities for learning:

“Well I think initially when this whole thing [QF] was being started some of the conversations were framed around what are the indicators of quality in a programme office that is doing well. How do we assess whether Ghana is better than Zimbabwe or vice versa. So I think the framing of the conversations around that time kind of planted the seeds of a league table…as long as people continue to see it [the QF] as a league table then we might see each other as competitors and therefore everybody [will keep] what he or she is doing very close to their chest” (Interview, Country Director 6).

Importantly, several features of the spreadsheet served to reinforce the ‘league table mentality’. First, it ‘was a series of numbers’ and did not contain any of the narrative discussion that featured prominently in the QF reports. Second, only element scores were displayed, with the indicator scores that led, at least in part, to the overall element scores being omitted. Third, each element score of 1 to 4 was assigned a colour (‘red and green and orange’) such that differences between countries in the spreadsheet were visually very distinct. ‘Low’ performers were particularly prominent as scores of ‘1’ were assigned the colour red. These features served to reinforce the focus on comparing countries, in effect privileging ‘competition’ over ‘learning’. This led regional directors to question the use of scores to promote learning, suggesting that it was their role, rather than that of a spreadsheet, to direct country directors to examples of good practice:

What’s the use of it [the spreadsheet]? What’s the use of sending it out to them [Country Directors]?...We [Regional Directors] are there to say ‘ok, actually, you know, you’re quite weak in this area, that country next door happens to do particularly well in this area, maybe we can put you in touch and you can share, learning, etc’, but putting a bunch of numbers…” (Regional Director 4, QF meeting)

More fundamentally, however, not only was comparison of countries in a spreadsheet not considered helpful for sharing best practice, but the uniqueness of each country also made such comparisons ‘unfair’:
Regional Director 1: I don’t see the value of knowing that, for example, on maybe even six of the twelve criteria, West Africa comes out worse than say South-East Asia because my interpretation instinctively would be what are the cultural, educational, historical background, you know, accumulation of circumstances in South-East Asia that means that they’re in a completely different environment.

Regional Director 4: It [a league table] makes it [comparisons] into a competition essentially.

PLA Staff 1: Yeah, but it’s an unfair competition...It’s like getting Marks & Spencers compared with Pete’s Café across the road where you’ve got totally different contexts.27

(QF meeting)

Claims of unfairness speak to tensions arising because the principle that country uniqueness is paramount was not being respected. Comparison of countries was reduced to a set of numbers, with little or no role for knowledge derived from either narrative accounts in the QF, or from regional directors’ own knowledge of the countries within their regions. The process of reducing performance on an element to a standardized metric was seen by country directors to have ensured that the contextual information required to understand these scores had been stripped away. Even when this information was present in the narrative section, it did not accompany the scores in the spreadsheet and was thus seemingly ignored (or considered too difficult to take into account). In this way, the ideals embodied in different modes of evaluation did not co-exist as the values of ‘competition’ had in effect ‘swamped’ the values of learning and country uniqueness. This was particularly problematic because improving quality through learning was an explicit ambition of the QF, with one of its stated purposes “to help identify strengths and areas for improvement across programmes so that steps can be taken to grow programme quality and foster learning” (emphasis in original).28

Much discussion centred on sharing examples of good practice amongst programme offices as a way to learn from others, but examples of this practice were rare. Country directors felt that there was no formal mechanism for sharing between programme offices, which was compounded by a lack of resources (both money and time). The view amongst country directors was that learning from the QF had been given a relatively low priority, and its

27 Marks and Spencers is a large UK department store whose annual revenue in the financial year 2009/10 was £9.5 billion. Pete’s Café was a small café immediately opposite the VSO building in London.

potential as a tool for sharing good practice went unfulfilled:

“What country directors want is to be able to go to a specific country where there’s good practice and drill down in that, you know, how did you do that? Why is it good? And they said that what we did last year didn’t give them that at all.” (Regional Director 5, QF meeting)

Reconciliation of the co-existing desires for learning and competition proved to be a central task in preparing the QF for its second year of operation.

Reordering priorities of learning and competition

Over a set of three meetings, staff in IPG debated the pros and cons of ranking programme offices using the QF data. At the first meeting, regional directors were generally against the idea of a spreadsheet being the vehicle for identifying good practice, instead arguing that they should take an active role in linking up programme offices to help improve performance. However, the IPG Director, one of the key proponents of using the QF to create competition, was not in favour of this approach and stalled any decision to the next meeting:

Regional Director 3: You don’t have to send it [the scores] out in a table to everybody but you can go and look and say ‘right, ok, this country over here does really good volunteer engagement, why don’t we arrange some sort of visit or some sort of support from that’, that would absolutely make sense but to send something out and say look for the [countries that have scored] fours and talk to them doesn’t.

Regional Director 2: Well what about if you sent out this summary…by element these are strongest programs…that would then be rewarding people who are doing well…what would people think about that…can we take a temperature check? I think we all agree there’s a reason to link people up according to where there is good practice, or good performance and there are ways to do that that aren’t about a published table. So, who’s for a published table, who’s against a published table, who’s for a published table at this stage?

[General laughs]

Regional Director 4: Obviously we [regional directors] are discouraging.

Director, IPG: I would reflect on it a bit further…I’m not sure how helpful it is when the table only had the element scoring and it’s very subjective…so I think work out if anyone else found it useful…I’m thinking all of this is worth picking up again in that next meeting.

(QF meeting)

Convened a few months later, the next meeting was focused on how to convince the IPG Director (not in attendance) that a ranking was not appropriate:

PLA Staff 2: Our task is to try and steer him (Director, IPG) in a slightly more acceptable direction.

PLA Staff 1: So what our current line is that we think it’s better for each country director to see their scores and then we use this process of analysis to highlight strong performances globally and preferable in each region so that you’d be able to say ‘on inclusion we think the countries that are stronger on this are these’, and then other CDs can contact them. So that gets us more away from that [a league table].
Regional Director 3: I’m just wondering what the reason for having a league table is, so that people will perform more effectively? And if that’s the end point then the route that you’re suggesting makes absolute sense because we can do that…I’m just wondering what lies behind that league table.

PLA Staff 1: I think it’s the idea you publish information and then people will be shamed, people will feel they got a low performance, they will feel forced to have to make improvement because it’s public.

PLA Staff 3: There is a real danger of labelling them [programme offices], isn’t there? That’s what’s really horrible about this because someone then gets labelled as being the office that’s rubbish at volunteer engagement or the one that’s great at such and such.

PLA Staff 1: Yeah, yeah I agree, yeah. The reason I really don’t like it, I don’t see how an organization’s [that’s] about volunteering and is very personal how that…sort of…philosophy could really fit with this, but the other thing is I think it will change the quality framework from being a learning tool… my real fear is if you publish the scores people get fixated on doing well on particular indicators, which we’re now saying aren’t good enough, rather than the spirit of trying to actually improve…so I think it’s a combination of philosophy in terms of what VSO is about but also, you know, keeping the quality framework as something that is a learning tool.”

(QF meeting)

In this brief interaction the key principles at stake in the compromise were explicated. The use of competition to ‘shame’ programme offices into improvements was viewed as not only ‘horrible’, but also fundamentally against the purpose of the QF as a learning tool, and even against VSO’s very philosophy as an organization. A month after this discussion, the same arguments were used in a meeting with the IPG Director, and a compromise agreed:

Director, IPG: Alright, let’s do it a different way, let’s ask each element leader, they’ve got to choose the top three [programme offices in each element]…let’s ask each element leader to highlight confidentially where they think there are real concerns.

PLA Staff 2: So, ok, that’s fine, then what? What happens to that information?

Director, IPG: So basically the element leaders are informing discussions about where we might prioritize…so the top three is highlighting good practice, giving an indication to countries across the world where they might want to talk to in terms of good practice, and the bottom one is just confidential for management purposes.

(QF meeting)

As such, the compromise between competition and learning, between comparisons and sharing good practice, was resolved through agreeing to differential disclosure. That is, the identity of good performers would be made public whereas the identity of poor performers would be kept confidential. Disclosure of good performers would allow the sharing of good practice between programme offices but without creating a ranking, and disclosure of poor performers to the IPG Director would allow management action to be taken but without ‘naming and shaming’ those programme offices in the process. It is here that debates about
the league table facilitated productive discussion between those who viewed ‘competition’ as
the route to improvement versus those who saw learning as the way to increase quality.
Unlike the debates over consistency in scoring, discussion was not focused on the QF per se,
but was connected to wider principles associated with particular world-views, particularly
those of a spirit of improvement, fairness and a volunteering ethos. Principled argument led to
a compromise between different world-views, despite strong enthusiasm for the spreadsheet
to remain. This ambition to share good practice was stated explicitly in the revised guidance:

“For sharing a list of those programmes that have performed very well in individual elements
with other programmes, it will be possible for CDs to directly contact their colleagues for
ideas or support in improving against areas of weakness. Over time these horizontal linkages
will help strengthen programme quality overall and will help develop direct programme
learning and peer support.”

29

However, the key to stabilizing the compromise was to implement the proposal effectively in
practice.

Embedding the compromise was most acute in the global analysis of the QF data.
Called an ‘Element Summary’, this analysis involved two persons, called ‘Element Leaders’,
being assigned to each element whose task was to review the element scores and narrative
across all programme offices and produce a summary analysis. Several changes were made to
the element summary process to help forge the compromise between learning and
competition. Unlike the first year where staff from IPG only completed the summaries,
Element Leaders in the second year involved country directors as well as staff outside of IPG.
This was viewed as bringing country directors back in to the process and involving them
directly in analysis and learning. The content and layout of the element summaries also
‘paired’ any analysis based on scores with that based on narrative reports. For example, the
following two questions were addressed, one following the other:

“What are the main global trends or anomalies in evidence from the statistical results for this
element?...

Based on the narrative from the country level reports, what are the main issues in terms of global programme quality in this element?” (emphasis in original).30

The identification of those elements with strong or weak performance also adopted a similar approach.31 For example, in relation to strong elements, the following areas were addressed, positioned in adjacent columns of a table:

“Top 5 areas of statistically improved programme performance i.e. those elements in which the highest number of POs improved their scores from 2007 – 2008...
...Top 5 areas of perceived progress i.e. those areas cited as an ‘area of progress’ during 2008 by POs” (emphasis in original).32

Here, analysis based on statistics was paired with that based on perceptions. But the identification and disclosure of good and poor performing countries still proved problematic. Consistent with the agreement, the spreadsheet of element scores was not distributed to country directors. Also, each element summary addressed the question “Where is good practice in the element occurring?” (emphasis in original)33, with element leaders listing countries that had performed well, along with specific examples of good practice. However, the competitive mantra and the desire to highlight poor performance lingered, with eleven of the fourteen element summaries identifying countries with “weak” performance, typically with reference to a “low” score on the element. Thus, despite intentions to enact a compromise between learning and competition, translating this fully into practice proved difficult.

Epilogue

Towards the end of the field study, a review of existing “Quality Initiatives at VSO” was conducted, as part of a move towards alignment of various practices into a single performance and accountability framework. The QF, along with six other initiatives, were

---

31 In 2009 country directors were required to list on the front page of the QF report “The 3 strongest elements in the PO”. These were not necessarily based on absolute element scores but on the judgement of the country director and other programme office staff.
33 QF Global Element analysis, 2009.
evaluated with positive and negative features outlined. Analysis of the QF was generally favourable but numerous “areas of improvement” were suggested:

Its holistic and coherent nature allows people to think more broadly and reflect on the progress of the whole programme…The process of doing the report makes people take stock, consider areas of improvement and make action plans accordingly….It seems that the QF is not referred to or used as often as people would like…The numbers are not useful because they are too mechanistic, yet subjective and inconsistent across POs… [Self-assessment] is a great way for the PO to take stock and think about their performance and how to make improvements. But many people feel that this needs some kind of external support and verification… although the design of the framework is quick and simple…It has become too long and its design means that indicators are ‘set in stone’ to a certain degree in order to make comparisons from one year to the next.34

This evaluation highlights how the compromises being made in the design and operation of the QF were not resolved but were on-going. The QF enabled broad thinking but was not used enough, self-assessment helped to ‘take stock’ but needed external verification, and the QF was simple, yet too long. We see that positive features of the QF that were closely aligned to one world-view inevitably gave rise to suggestions for improvement that sought to address the concerns of those with different preferences. In this way, the co-existence of and reconciliation between different worlds can be seen as an enduring feature of a compromising account.

Discussion and conclusion

Our study is important for analyzing an unexplored role of accounting in organizational change. As noted previously, prior research has shown how an existing organizational culture was displaced by another and in the process, new or adapted types of accounting were developed to support, or even constitute, such change (e.g., Dent, 1991; Oakes et al., 1998; Ezzamel et al., 2008). In contrast, our study highlights the role of accounting in achieving co-existence and reconciliation between different ‘worlds’ or cultures, rather than the use of accounts in processes of conflict and displacement. A focus on the co-existence of different modes of evaluation and attempts at their reconciliation provides a lens through which to

examine more subtle (but no less important) shifts in organizations, particularly where change involves an ongoing co-existence of and ‘rebalancing’ between competing values rather than shifts from one to another. The trend towards an increasing plurality of interests in organizations makes an analysis of the role of accounting in compromising between co-existing world-views particularly important.

**Co-existing worlds and modes of evaluation**

Accounts are important in compromising between different world-views because they are sites where multiple modes of evaluation all potentially operate at once (Stark, 2009). VSO’s attempts to measure the performance of its programme offices brought together differing modes of evaluation, one based primarily on ‘Learning and Uniqueness’ and the other based primarily on ‘Consistency and Competition’. These modes of evaluation had their origins in the privileging of different ‘worlds’ and their associated value systems, where each mode of evaluation can be distinguished according to its purpose, the desirable attributes of a good evaluation and subsequently the desirable attributes of a good account (see Table 3 for a summary of these two modes of evaluation and their characteristics). Our field study shows that it is in routine choices about issues such as indicators, scoring processes, the identification of good and poor performers, or methods of data analysis, that compromises between these differing modes of evaluation and orders of worth were made.

The QF was subject to on-going criticism and was ‘loved by no-one’. Rather than presenting an obstacle to success, we suggest that it is the ‘imperfect’ nature of the QF that was pivotal to its continued existence as a compromising account. We see that the constant shifting and rebalancing in the design and operation of the QF enabled the co-existence, albeit often temporary, of different modes of evaluation. Changes that privileged one mode of evaluation, such as a focus on a more rigorous and consistent scoring process, were
accompanied by changes that shifted the emphasis back to another mode of evaluation, such as ensuring the analysis of QF data included a pairing of numbers with narrative. It was this constant rebalancing and ‘give and take’ between different modes that helped to resist pressures for recourse to a single and therefore ultimately dominant mode of evaluation (c.f., Thévenot, 2001). In contrast, the privileging of particular modes of evaluation to the exclusion of others in the SRA and ACR, meant these two approaches were ultimately discarded.

We tentatively suggest that an important feature of a compromising account is that of ‘concurrent visibility’. In reconciling different modes of evaluation, it was critical that the QF made visible the features of an account that were important to different groups. Here we use the term ‘visible’ in a broad sense to refer to how the design and operation of a compromising account reveals the attributes of accounts that are important to different groups. For example, in the physical format of the QF indicators were accompanied by narrative boxes, and in the analysis of the QF data, ‘statistical’ analysis was accompanied by analysis based on ‘perceptions’. The concurrent use of these different features gave explicit visibility to the importance of different modes of evaluation and their world views. In contrast, only a single mode of evaluation was made visible in both the SRA (namely, scoring) and the ACR (narrative) such that particular world views were left invisible. We suggest that where the co-existence of different world views is an on-going feature of organizational life, groups are likely to be particularly concerned that fundamental principles from their world may not be respected and thus come to be dominated by others (c.f., Denis et al., 2007). It is here that ‘concurrent visibility’ in a compromising account can provide confirmation and reassurance that a particular mode of evaluation and associated world view is, indeed, recognized and respected, thus making ongoing reconciliation and compromise more likely.

An examination of compromising accounts also reveals how the account itself can act as
an agent in the ongoing reconciliation between different divergent ‘worlds’. A compromising account is thus both the process of, and at particular moments the specific outcome of, a temporary reconciliation between competing modes of evaluation and orders of worth. Analogous to Chua’s (2007) discussion of strategizing and accounting, this draws attention to a compromising account as both a noun, i.e., the account itself that is produced in some material form (e.g., a balanced scorecard, a financial report), and as a verb, i.e., the acts that comprise the process of reconciliation that lead to and follow on from the physical production of an account. In this way, accounts, such as VSO’s QF, not only reflect particular combinations of differing modes of evaluation, but also create opportunities for the bringing together and thus potential ‘re-ordering’ of competing value systems. A compromising account can thus be a vehicle through which dialogue, debate and reconciliation between groups is produced, even if only temporary. It is the productive discussion that can come forth from having to compromise on the account that can be beneficial.

**Processes of reconciliation**

But compromises are not always readily forthcoming and of necessity reflect temporary agreement between groups and their respective worlds. This is because the compromise, whilst enabling agreement, suspends competition between worlds rather than resolves it (Thévenot, 2001). We see that breaking points in the compromising account emerged particularly in debates over ‘incommensurables’ where particular groups felt that fundamental principles associated with their world were not being respected (Espeland & Stevens, 1998; Denis et al., 2007). This was most acute in discussions surrounding the scoring of indicators and elements where changes to the process privileged the value of ‘consistency’ above that of country ‘uniqueness’. In addition, the presentation of scores in a spreadsheet and the subsequent illusion of a league table privileged the value of ‘competition’ above that of
'learning’. In both cases, a group that felt that a fundamental principle was not being respected and countered the existing practice by denying that the performance and hence value of countries was comparable because of their ‘unique’ contexts.

We see that claims of incommensurability were not mere technicalities (although they were sometimes expressed as such) but can be viewed as “vital expressions of core values” (Espeland & Stevens, 1998; 327) intrinsic to a particular mode of evaluation and order of worth. Debate over the commensurability of items to be included in a compromising account, manifest in choices about the use of particular metrics, images and words, were likely to bring different modes of evaluation into conflict. The response to these moments was critical to ongoing viability of the compromising account. In the case of the league table, the spreadsheet was not distributed, but countries performing well or poorly were still ‘named’. Hence, we see movement towards a ‘rebalancing’ of interests that helped, to some extent, to restore a sense of co-existence between the different modes of evaluation. In the case of scoring, the push for consistency and standardization resulted in strong criticisms that the QF was not a useful vehicle for learning at all, leaving the QF in a tenuous position at the end of its second year of operation.

The marshalling of particular styles of argumentation in reconciling different world views was also critical. Drawing on Stark (2009), we differentiate between ‘personalized’ and ‘principled’ forms of disagreement that are used in situations where differing modes of evaluation co-exist. Personalized disagreement involved recourse to arguments that invoked personal ‘preferences’ to support a particular position. For example, concern from country directors over the use of particular indicators was often supported with reference to a dislike of ‘numbers’, ‘indicators’ or ‘scores’, and a fondness for narrative. Personalized disagreements tend to endure without resolution because they are typically dismissed as reflecting individual ‘preferences’. For example, these arguments from country directors
about indicators were not acted upon as they were viewed as (mere) ‘complaints’. In these situations, arguments were on-going, but led to inaction and a lack of accomplishment (Stark, 2009). In particular, much debate was focused on the design of the account itself, rather than it serving as a mechanism to debate the different world-views in play.

In contrast, principled disagreement involved recourse to arguments that invoked a particular world view to support a position. This was most evident in the ‘league table’ debates, where a ‘spirit of improvement’ and the nature of VSO as an ‘organization’s about volunteering’ were used in arguments against the distribution of scores in a spreadsheet. Principled disagreements are more likely to lead to a (re)consideration of the orders of worth because they are typically viewed as recognizing the broader principles at stake in the compromise between differing modes of evaluation. For example, these arguments against the ‘league table’ led to a change in practice whereby values of ‘learning’ and ‘uniqueness’ were given a more equal footing with the (previously dominant) emphasis on ‘competition’. As such, principled disagreements are likely to involve shifts in practice such that on-going reconciliation and co-existence of differing modes of evaluation is more forthcoming.

**Developing performance measurement systems**

Our study also has implications for research on the development of performance measurement systems. Much of this research examines ‘design’ issues, for example, the mix between financial and non-financial metrics (e.g., Cardinaels & van Veen-Dirks, 2010; Baines & Langfield-Smith, 2003), or how to validate cause-and-effect links between different performance measures (Malina et al., 2007). These issues are (often implicitly) analyzed as technical problems, such that their solution lies in the application of expertise in performance measurement system design. Whilst clearly important, our field study shows that the issue of what is considered an appropriate ‘balance’ between indicators or narrative, or the correct
‘weight’ of different elements, is fundamentally connected to the beliefs of different groups regarding the attributes of a ‘good’ account.

We see such normative beliefs in the rhetoric surrounding the promotion of the balanced scorecard, which makes explicit claims as to what constitutes a ‘good’ account of performance. For example, a preference for greater inclusion of non-financial measures is because the “traditional financial performance measures [that] worked well for the industrial era… are out of step with the skills and competencies companies are trying to master today”. And, a focus on financial metrics is explicitly associated with the “engineering mentality of the industrial age” and not with the “kind of organization many companies are trying to become” (Kaplan & Norton, 1992: 71, 79). These statements have an explicit normative dimension that privileges particular types of metrics (non-financial over financial) and thus particular modes of evaluation (‘information age’ over ‘engineering mentality’). Here design issues are not merely technical phenomena but represent a broader struggle between the attributes of accounts that are deemed important under differing modes of evaluation. In this way, our study indicates that insight into the development of performance measurement systems, like the balanced scorecard, can be furthered through explicit examination of how choices about particular metrics, weights and presentation formats are linked to different modes of evaluation and their attendant world-views.

Our study also shows how values as well as knowledge are critical to the development of performance measurement systems. Wouters and Wilderom (2008) argue that the failure of many top-down performance measurement initiatives lies in the insufficient use of the local experience and unique expertise of employees. Here, developing better performance measures means taking “advantage of local knowledge” (Wouters & Wilderom, 2008: 493). Although integration of local knowledge is important, our study indicates that the development of performance measures is also a process of reconciling differing values, particularly as the
choices made in developing a performance measurement system act to validate some values and discredit others (c.f., Stark, 2009). In this way, attention to local experience is broader than merely recruiting the best available knowledge; it is also concerned with the recognition and incorporation of the different values and modes of evaluation that are privileged at different organizational (e.g., local vs. top management) levels.

Our study has implications for how to study the role of MCS in organizational change. The concept of a comprising account as one that can assist change by reconciling different worlds has been useful in studying changes in a performance measurement system. Such an approach could be applied more generally by researchers to study other aspects of MCS that involve both qualitative and quantitative approaches (such as balanced scorecards or budgets) in facilitating change in other sectors including both public and private organizations. The study finds that the use of performance measurement systems will be influenced by how components of the system relate to potentially competing world-views and how they interact through time. This approach is particularly pertinent to researchers wishing to study how change initiatives can lead to acceptance and use of MCS. The focus on compromising accounts provides a rich approach to understanding how different aspects of performance measures change in response to competing world-views and provide a stage for these views to be contested through time. It is the processes by which the performance measures and world-views interact that provide researchers with a way of examining the likelihood of effective change.
References


Table 1
Formal fieldwork activity

<table>
<thead>
<tr>
<th>Interviews</th>
<th>Location of staff</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Director, International Programmes Group</td>
<td>London</td>
<td>2</td>
</tr>
<tr>
<td>Deputy-Director, International Programmes Group</td>
<td>London</td>
<td>1</td>
</tr>
<tr>
<td>Regional Director</td>
<td>London, Ghana</td>
<td>2</td>
</tr>
<tr>
<td>Head-Programme Learning and Advocacy</td>
<td>London</td>
<td>1</td>
</tr>
<tr>
<td>Team Leader-Programme Development and Learning</td>
<td>London</td>
<td>2</td>
</tr>
<tr>
<td>Executive Assistant to Director, International Programmes Group</td>
<td>London</td>
<td>3</td>
</tr>
<tr>
<td>Programme Learning Advisor</td>
<td>Ottawa</td>
<td>1</td>
</tr>
<tr>
<td>Systems and Project Manager</td>
<td>London</td>
<td>1</td>
</tr>
<tr>
<td>Head-Strategy, Performance and Governance</td>
<td>London</td>
<td>1</td>
</tr>
<tr>
<td>Director-VSO Federation</td>
<td>London</td>
<td>1</td>
</tr>
<tr>
<td>Volunteer Placement Advisor</td>
<td>London</td>
<td>1</td>
</tr>
<tr>
<td>Finance Manager</td>
<td>Sri Lanka</td>
<td>1</td>
</tr>
<tr>
<td>Programme Manager</td>
<td>Sri Lanka</td>
<td>2</td>
</tr>
<tr>
<td>Facilities and Office Manager</td>
<td>Sri Lanka</td>
<td>1</td>
</tr>
<tr>
<td>Volunteer</td>
<td>Sri Lanka</td>
<td>2</td>
</tr>
</tbody>
</table>

32

<table>
<thead>
<tr>
<th>Observation and attendance at meetings</th>
<th>Location of meeting</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality Framework meetings</td>
<td>London</td>
<td>6</td>
</tr>
<tr>
<td>Various planning and review meetings</td>
<td>London</td>
<td>6</td>
</tr>
<tr>
<td>Programme planning and review workshop</td>
<td>Sri Lanka</td>
<td>3</td>
</tr>
<tr>
<td>Office planning and logistics meeting</td>
<td>Sri Lanka</td>
<td>2</td>
</tr>
</tbody>
</table>

17
Table 2  
Quality Framework indicators: change to indicator 8.1 between 2008 and 2009

<table>
<thead>
<tr>
<th>Year</th>
<th>Key performance indicator description</th>
<th>Performance range guide</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>Number of proposals signed off by the CD/RPM against agreed quality criteria in Stage 2 of PMPG</td>
<td>Performance to be assessed on a country by country basis</td>
</tr>
<tr>
<td>2009</td>
<td>Value of proposals signed off by the CD/RPM against agreed quality criteria in Stage 2 of PMPG</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 More than £550,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 £350,000 - £550,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 £150,000 - £350,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 Room for development Less than £150,000</td>
</tr>
</tbody>
</table>
Table 3
Modes of evaluation at VSO

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Modes of evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“Learning and Uniqueness”</td>
</tr>
<tr>
<td></td>
<td>“Consistency and Competition”</td>
</tr>
<tr>
<td>Purpose of evaluation</td>
<td>Reflection, learning, improvement</td>
</tr>
<tr>
<td></td>
<td>Standardize, compare, compete</td>
</tr>
<tr>
<td>Attributes of ‘good’ evaluation</td>
<td>Contextual, detailed, 'local' interpretations</td>
</tr>
<tr>
<td></td>
<td>Consistent, precise, objective, ‘universal’ interpretations</td>
</tr>
<tr>
<td>Attributes of ‘good’ accounts</td>
<td>Narrative descriptions, case studies, stories, images</td>
</tr>
<tr>
<td></td>
<td>Numbers, indicators, and scales, particularly those that can be compared between units</td>
</tr>
<tr>
<td></td>
<td>Indicators that capture current performance accurately</td>
</tr>
<tr>
<td></td>
<td>Avoid reliance on narrative as it ‘selective’ and cannot be compared between units</td>
</tr>
<tr>
<td></td>
<td>Avoid reliance on numbers as they provide only a partial account and do not tell the ‘real’ story</td>
</tr>
</tbody>
</table>
Figure 1
VSO Structure

Council
(22 Members)

International
Board of Trustees
(13 Members)

Senior
Management Team
(6 members)

Federation Members
United Kingdom
Netherlands
Canada
Kenya
Philippines

Other recruitment bases
Ireland
India
Australia

International
Programmes Group
Director and Senior
Managers
Programme Learning and
Advocacy Group
Programme Funding and
Development Group

Regions
Regional Directors

Country Programmes

Programme Office
Country Directors
Programme Managers
Finance, Administrative
and Logistics staff

Partner Organizations
Volunteers
Appendix 1: Strategic Resource Allocation tool

Assessment of Programme Effectiveness
Country: …………………………………

The higher the overall percentage a Programme Office receives in this tool, the more “effective” it will be perceived to be based on this measure.

Section A. Focus on disadvantage (48 % of total score)

<table>
<thead>
<tr>
<th>Measure</th>
<th>% of total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 HDI</td>
<td>17%</td>
</tr>
<tr>
<td>2 Percentage of more disadvantaged people being reached through implementation of CSP aim</td>
<td>10%</td>
</tr>
<tr>
<td>3 Scored analysis of how well strategies are working in addressing the causes of disadvantage</td>
<td>10%</td>
</tr>
<tr>
<td>4 Disadvantage Focus in Current and Planned Placements</td>
<td>11%</td>
</tr>
</tbody>
</table>

Section B. Outputs of Country Programme (27% of total score)

<table>
<thead>
<tr>
<th>Measure</th>
<th>% of total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 What % of placements in the last 2 planning years have fully or mostly met their objectives (not including early return reports)?</td>
<td>13%</td>
</tr>
<tr>
<td>6 What was the Early Return rate (excluding medical &amp; compassionate) over the last two planning years?</td>
<td>4%</td>
</tr>
<tr>
<td>7 What percentage of the ACP target of fully documented requests (i.e. with Placement Descriptions) was submitted on time over the last 3 planning years?</td>
<td>5%</td>
</tr>
<tr>
<td>8 What percentage of the ACP target number of volunteers was in country on 31/3/01, 31/3/00 &amp; 31/3/99?</td>
<td>5%</td>
</tr>
</tbody>
</table>

Section C strategic approach (25% of total)

Note that the statements attached to each score are for guidance and are not absolute statements: we recognise that with some programmes no one statement will accurately describe the programme. The RPM must have a clear idea of the rationale behind the scoring, in order to ensure transparency and to allow comparison between countries. All of your scores should be based on an analysis of the current situation – i.e. not future strategy or placements.

<table>
<thead>
<tr>
<th>9. Strategic approach based on programme at the current time</th>
<th>Score</th>
<th>% of total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Placements working at different levels (micro/macro) towards strategic aims + planned links between them.</td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>b) Critical appraisal of placements with clear rationale linking placement to strategic aim + planned exit strategy</td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>c) Strategic &amp; linked implementation of cross cutting themes</td>
<td>2%</td>
<td></td>
</tr>
<tr>
<td>d) In-country advocacy by the programme office</td>
<td>2%</td>
<td></td>
</tr>
<tr>
<td>e) PO proactive in promoting increased development understanding amongst volunteers</td>
<td>2%</td>
<td></td>
</tr>
<tr>
<td>f) Openness and commitment to learning</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>g) Genuine partnership relationship with employers and other development actors</td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>h) Types of placements used most appropriate to needs of disadvantaged groups and based on strategic reasoning.</td>
<td>2%</td>
<td></td>
</tr>
<tr>
<td>Name of Programme:</td>
<td>Please enter name of country here</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Element</strong></td>
<td><strong>Indicator</strong></td>
<td><strong>Indicator result</strong></td>
</tr>
<tr>
<td>Programme outcomes at partner level</td>
<td>A.1 Annual Progress in PAP objectives is achieved</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>A.2 Programmes and partners monitor and review progress in capacity development and/or service delivery</td>
<td>4</td>
</tr>
<tr>
<td>Programme impact at beneficiary level</td>
<td>B.1 Positive changes for target groups of partners are achieved</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>B.2 Programmes and partners are able to understand, monitor and review changes for target groups</td>
<td>4</td>
</tr>
<tr>
<td>Relevant and ambitious strategic plans are evolved in response to the development needs of the country’s disadvantaged communities</td>
<td>1.1 PO is responsive to changes in social, economic and political context</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1.2 PO is consulted by peer agencies and/or government bodies a credible development agency within its field of operation</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1.3 Programmes working at different levels (e.g. national, provincial, districts, grass-roots) towards strategic aims</td>
<td>4</td>
</tr>
<tr>
<td>Appropriate and innovative use of development interventions to deliver programme outcomes and impact</td>
<td>2.1 The contribution of National Volunteering (NV) to programme delivery has been maximised</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2.2 The contribution of a range of different interventions to programme delivery has been maximised</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2.3 Development awareness amongst volunteers and the wider community has been maximised</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2.4 Opportunities to develop international resource partnerships are fully explored and developed</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2.5 The contribution of advocacy to programme delivery has been maximised</td>
<td>4</td>
</tr>
<tr>
<td>Programme delivery against plans</td>
<td>3.1 LTV/YPD &amp; STV arrivals during year against reforecast plans</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>3.2 Firm and documented placement delivery against reforecast plans</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>3.3 Quality of placement documentation</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>3.4 PIP milestones successfully completed across all programmes</td>
<td>4</td>
</tr>
<tr>
<td>Inclusion</td>
<td>4.1 The PO has an inclusion statement and standards that are shared among all staff and that new staff sign. This is for both programme work and how the programme office is run</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>4.2 Partner organisations include excluded groups in their work and as part of their target group</td>
<td>4</td>
</tr>
<tr>
<td>Planning and review</td>
<td>5.1 Number of PIPs updated and signed off annually as a result of PAs in line with guidance</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5.2 All programmes are reviewed annually in line with guidance</td>
<td>4</td>
</tr>
<tr>
<td>Partnership development and maintenance</td>
<td>6.1 Portfolio of partners in place relevant to PAP and CSP objectives</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>6.2 Long-term (3-5 years) Partnership Plans are in place which include partnership objectives that are linked to the PAP objectives</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>6.3 Partners are actively involved in programme development and review</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>6.4 Partnerships are reviewed annually to assess progress towards Partnership and PAP objectives and quality of the relationship with VSO</td>
<td>4</td>
</tr>
<tr>
<td>Volunteer engagement and support</td>
<td>7.1 Volunteer support baselines are being met by the Programme Office and partners are supported to manage volunteers</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>7.2 PO celebrates volunteer achievement, responds to volunteer problems effectively and encourages the development of effective and accountable volunteer groups</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>7.3 Volunteers are engaged in programme development</td>
<td>4</td>
</tr>
<tr>
<td>Programme funding</td>
<td>8.1 Value of proposals signed off by the CD/RPM against agreed quality criteria in Stage 2 of PMPG</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>8.2 Restricted income as % of PO Total Budget (including vol. recruitment costs)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>8.3 Donor conditions for existing funding have been met (including financial, narrative and audit reports submitted on time and to the standard required by the donor) throughout the year</td>
<td>4</td>
</tr>
<tr>
<td>Cost effectiveness</td>
<td>9.1 Percentage of total approved PO managed budget (restricted and unrestricted) budgeted on PC and VC costs in 2008/09 (Global average = 46%, Regional averages range from 32% and 57%)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>9.2 Possible areas of saving against costs identified during budget setting process through innovation and creative thinking</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>9.3 Percentage of total revised PO managed budget (restricted and unrestricted) budgeted on staff costs in 2008/09</td>
<td>4</td>
</tr>
<tr>
<td>Financial management</td>
<td>10.1 Annual programme office expenditure variance (restricted plus unrestricted) for 08/09 against budget adjusted for macro-forecast</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>10.2 Volunteer Unit Cost based on 08/09 budget</td>
<td>4</td>
</tr>
<tr>
<td>Staff management and support</td>
<td>11.1 Performance management system is being actively implemented</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>11.2 Evidence of major HR policies and systems being adhered to</td>
<td>4</td>
</tr>
<tr>
<td>Legal and policy compliance and risk management</td>
<td>12.1 Number of outstanding category A and B internal audit actions for PO action relating to legal compliance</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>12.2 Security Risk management plans signed off and implemented and tested according to Country’s main security risks (e.g. avian flu, security, natural disasters etc)</td>
<td>4</td>
</tr>
</tbody>
</table>