International Conference on
Human Resource Development in the Public Sector
Quality Training & Excellent Performance

捌、第三場次專題研討：
績效指標的運用與限制

Session III: Utilization and Limitation of
Performance Indicator

捌、第三場次專題研討

一、時　間：101 年 10 月 16 日 (星期二) 下午 1 時 50 分

二、議　題：績效指標的運用與限制

三、主持人：黃俊英 考試委員

四、演講人：John Wanna 澳洲國立大學政治與國際關係學院教授

五、與談人：Peter Vandenbruaene 比利時聯邦公職訓練所人力訓練所人力訓練處處長

　　　　　　史美強 東海大學公共行政學系教授

　　　　　　張四明 國立臺北大學公共行政暨政策學系教授

六、主持人：

　　伍副院長、各位貴賓、各位女士、各位先生大家午安，我是黃俊英，非常榮幸主持這場專題研討，我們要談的是績效指標的運用與限制，很榮幸邀請 John Wanna 教授，Wanna 教授目前任教於澳洲紐西蘭政府學院，另有三位與談人 :Peter Vandenbruaene 來自比利時、及兩位臺灣的學者張四明教授，以及史美強教授。我們首先歡迎 Wanna 教授。

七、演講內容：

　　謝謝黃委員，各位午安，非常榮幸能夠參與本次會議，謝謝臺灣主辦單位對我們的熱情款待及經驗分享，我在這兩天聽到全球的意見交流及經驗分享並將其帶到臺灣來。我今天要做的是提醒大家思考交流的意義為何，我們不應盲從的追隨別人，而是應該要質疑其他國家做事方法，並進行測試，而不是把一套方法原封不動的套用而已，我們不應該去誤解其他國家做事的方法，而是應該從別國的錯誤中學習，不要重蹈覆徹，我在準備今天的講稿時，好像是屬於比較批判的聲音，不過在這兩天的活動中，我不斷聽到大家的回應指出大環境的重要性，而不是盲目追求績效的指標，另外也有人提到學習的困難，且可以將什麼樣的東西帶到臺灣，而今早 Matthew 也有提到我們應該要清楚地思考，學習的角色是什麼，學習的內容是什麼，而葉次長也提到公私部門之間的結合，如何互相學習。我必須先跟大家強調我並不是反對績效評估，我只是認為很多時候我們進行方法並不是最正確的，或是我們誤解了績效評估，反而讓績效評估限制了我們的選擇，而不是提供我們更多的選擇，所以應該先理解大環境的脈絡為何。我認為績效評估往往被誤用，這個評估往往限制了我們的行動，而我的演講也旨在說明績效評估在公部門間遭受到的濫用和誤用。

　　昨天，我們有討論到臺大的 EMBA 評比好像下滑了，也講到了相關的機會何在？但我覺得在績效管理方面，像臺灣、新加坡或香港這樣比較小的國家，事實上，它們可以扮演領導世界的角色，比起其他很多大國在績效管理方面可以做得更好，現在我要進入演講的主題。

　　我們現在碰到什麼樣的挑戰呢？請大家先退一步想，我認為現今的公部門不但花了龐大的經費也面臨了許多的限制，而公部門也希望所投入的經費可以看到收穫及成果，但我們並不清楚投入與結果之間的關聯在那裡，我們會致力推廣政策也希望透過績效管理讓我們的媒體、政治人物及一般人看到，是否有達到我們所預期的目標。但我們也應思考目標到底是什麼？很多人會覺得所謂的績

效管理就好像是特效藥一樣，意指大家將績效管理當做是一種神聖的指標或最高指導原則一般，但或許它其實是無形的、是不存在的，好像只是個理想而已，大家一直盲從的追求績效管理，把它當做是聖杯一般。大多數人都覺得績效評估是一件好事，當在教科書學習的時候，覺得它是一件好工具，但並不清楚其機制為何，似乎每個人都以為只有古代人才不喜歡所謂的"績效評估"，不喜歡或不接受它的人也遭受到社會很大的批判，但績效管理或績效評估這樣的概念是大家從未去質疑的，所以我覺得要進一步的去了解它、質疑它，就我本身來說，並沒有親自在公部門當中實施績效評估的指標，告訴你說哪些制度有用、哪些沒用，但我在別的部門評估公共政策，還有在財務機構或者是某些部會思考他們到底提出什麼樣的牛肉或是什麼樣的料？那我認為我們這邊有假設、執行及結果，然而，大家過於相信績效評估，認為這個體制太過完善，這就是我們必須深思熟慮的問題，除此之外，我們也覺得所選擇的績效指標能提供有效的資訊，事實上，它的確提供了資訊，但試想之，這些資訊真的是有用的嗎？是可以在公部門妥適運用的嗎？另外，績效管理的體制本身或許是有缺陷的，因為我們所選擇的指標是不良的，導致我們無法推展良好的行為，且我們常看到的是外部容易顯現的結果，而內部所隱藏的缺失是我們看不到的，以及我們對於某部份採績效評估或績效指標的單位是不信任的，那很多時候這會是個問題，那目前我們的中央化、一致化或標準化的關係，這也是我們的問題之一；另外，以績效評估結果面來說，這些結論會是我們判斷某些決策很好的依據，但事實上，真的是如此嗎？我們真的能確切了解到績效評估所代表的意義是什麼嗎？譬如說談論到某些學校的績效比較不好或者某些學校地處偏遠，在我們操作績效評估所得到資訊，是否能夠正確解讀呢？而且，我們本身對績效評估解讀的能力不足，導致我們無法將績效評估與政府機制做結合。除此之外，我們常用績效評估這個名目來編列預算，大家會容易誤認它是一個很有革命性的作為或者改革，雖然，這樣的想法能吸引大眾注意，但不見得是正確的作法，那到底進行績效管理的目的為何呢？大家認為答案是最顯而易見的。提倡此作法的人認為有以下的原因是為了要改善體制中的某些問題或者是政府整體等；除此之外，透過績效管理我們可以介入某些政策，譬如說大家認為績效管理就是政府某些行為背後的原因，它也可以是良好政策規劃的元素，譬如說我們要執行某個計劃，想知道它是否有效？是否可以從中獲得一些好的結果；除此之外，績效管理也可以在經營單位中當成持續進步的元素，績效管理也可被政府用來競爭力的評比，譬如說機關（構）比機關（構）之間表現較好；政策與政策之間表現更好，績效管理亦可作為保護公部門的防禦機制，但我個人認為這會引起外界的質疑。我覺得某些評估制度會有些改變，它會有些融合，而不同的制度也會有不同結果和看法，很多時候，績效評估是掌握在他人手中，流於過度官僚而無法貼近機關（構），我們所面臨的問題並不完全是我們自己造成的，政府會去評估所有的方案及計劃，那這時候績效評估的結果是不務實也是不切實際的。然而，一直不斷進行績效評估並不是最好的事情，這也會演變成機關鬥爭的工具，會剝奪或增強某些機關的權力。除此之外，不同單位會有不同的目標，譬如說我們可能會認為某些單位運作的不錯，但事實上並非如此；也會針對立法進行評估，或是為一般大眾或媒體來進行績效評估，媒體通常對績效評估不感興趣，只有在表現不好時會吸引他們的注意力；再者，也會對其它行業進行所謂的績效評估，像醫師或其它行業。其實績效評估也會受到不同思潮或科技影響，但這變得只看到可量化的結果，就無法看到質化的問題，當我們談到績效評估指標的時候，之前葉次長就有提到類似的看法，其實它就是做什麼事就會評估什麼結果，亦即我們所評估的指標就是我們會做的事情，這樣我們就無法注意到事情真正的問題所在，我們常以量化來評比一個單位的表現，譬如說我們要評比學校的表現，就要看它師生的比例或聘請多少老師等等，但這就無法看到真正教育的品質，我們只能知道學校是否有符合規定，但政府也想知道當政策在執行時會產生什

麼樣的結果，且我們可以解決什麼樣的問題，但無法看到指標和結果之間的關聯性。以下是我們常用的三個常用指標：1. 業務量－即我們花了多少時間和力氣去完成這些事；2. 過程數據－即我們處理多少案件或業務量到底有多少等等，是否有符合民眾的需求；3. 投入與結果的比例－即投入的成本或經常的成本等，實際上我們都無法精確的掌握結果，因為我們只能得到量化的資料無法得質化的結果，這很容易引起爭議，這些指標都只顯示我們投入的努力，這也是績效評估的矛盾。其實績效評估要做得好必須要去評量的應該是看不見的東西。再者，由於大家都需要處理很多的表格，當我們在評估一所醫院或學校時，大家都想要用表格來統整，其餘如環境或背景這些很難以放進表格的文字敘述是很難評比的，因此，導致在評量時大家都以數字導向的方法來參與評比，這樣是很容易誤導他人的。例如十年英國政府就是太著重於數字評比，於是許多人都扭曲了遊戲規則，因為有了規則，所以他們就想辦法鑽漏洞，反而沒有達成指標想達成的效果。除此之外，這些績效指標都是相互關聯的，在績效評估的過程中，是有不同的階段，其中有很多出錯的地方，而這也會和辦公室文化有關聯，首先，先蒐集資訊，再經由資訊產出，然後再衡量績效，亦即統整資料，最後，再做出績效報表，我們報告時，當然只寫我們認為沒有問題的東西，再者績效會有一些評鑑或稽核，但稽核人員做的工作是相對較低的，也有一些是個案研究，個案研究是可以提供一些背景以及大環境的資訊，除此之外，我們再做一些績效評估，這裡就有一些判斷的元素，還有就是績效的預算，就是把績效及預算做結合，另外還有就是全面性的績效管理，它需要前面幾個階段都做得很好，但前面的階段都有可能出錯，都有可能造成最後的資訊是被誤解的，或是被不當加油添醋，再來，績效指標還有一些問題，首先就是政府，由於蒐集到的資訊並非最重要的資訊，我們在澳洲有一些專案所談論的是社會和諧，社會和諧要蒐集什麼資訊呢？是要了解族群之間有什麼衝突，發生的次數，若是衝突，我們可以放衝突的次數，如果是社會和諧或族群和諧的話，我們要放什麼要的指標呢？我們並沒有好的指標，再者我們也沒蒐集這些資訊的良好方法，譬如說我們想要有更好的社區亦即夜晚走在路上不會被攻擊等等，這類的情形你無法衡量，也難以彙整資料的；再來，專案部份，它其實是由地方到中央政府的資料傳輸，所以在其中有牽涉到不同層級的政府，且可能各有不同的目標及監督方式，所以這樣出來的結果，也是容易遭人質疑的，再來，我們有太多不相關的指標且大家可能會愈加愈多的指標，譬如說避免歧視的指標之類，再做績效指標時，只有指標愈加愈多，並沒有把不相干的指標剔除，有關不相關指標部份，譬如說監獄房間大小、床的大小等等，這類的指標會不斷的增加，卻沒有將不相關的指標除掉，導致我們有指標愈來愈多的問題，其實正式的施政目標及非正式的政治目標這兩者是有差異的，譬如說正式的施政目標是要做某些事情，但非正式的政治目標是希望媒體報導的都是好的事情，這樣政治人物在選舉的時候就會有好的表現等，所以，大部份的時候都是這種非正式的政治目標主導，再者，有時候我們會為了資料而追逐資料，但這些資料可能不是所謂重要的實務資料，譬如說都會區的貧窮率，我們要如何蒐集呢？會牽涉什麼樣的族群呢？這必須問實際在做事的經理人是最清楚的，但現在我們往往為了追資料而追資料，再者，現在大部份的 OECD 國家並沒有獨立第三方的人來監督績效指標，可能是政府不願意或資源不足，但不管怎樣，就是沒有人來監督說這個資訊是否正確，除此之外，績效指標有很多衝突的層面，譬如說效率與效度這兩者是衝突的，還有公平問題及個別需求的問題，這其實不同是指不同指標之間是有衝突的，特別牽涉到公部門所重視的社會責任的時候。

接下來，我們來談績效管理的迷思，很多公部門的人認為當我們有績效評估時，民眾就會對公部門的信任度提昇，但其實不然，所謂的信任是指政府的體制而不是所謂的績效管理，但還是會有

人說社會大眾或政府部門很重視這一點，但事實上是不然，社會大眾只有醜聞出現時，才會關注績效管理部分。所以有 96% 的時間，你的報告是沒有人讀的。再來，有人認爲若有績效管理制度的話就會有更多的創新，這也是個迷思，因爲有了績效管理會讓人更不敢做事；但也有人認爲有績效管理會有更多的效率，因爲有提供績效獎金就可以提高績效，有些國家政府是有提供這樣的做法，比如說有提供 30%-50% 的績效獎金，但並沒有達到所謂激勵的效果，因爲有的人如果沒有拿到獎金，就不會朝這方面努力；再者還有另一個迷思，就是可以激勵服務的提供者，但要激勵別人是很複雜的事情，激勵的動機可能來自他目前的職位或他追求的目標，不是說他的績效評分裡打了勾，他就很高興，就得到激勵。再來，有人說我們現在的政策應以實證爲基礎，但現在並不是如此，我們有許多政策不是根據證據或是實證在說話。再者，評估績效的問題，首先是因果關係並不清楚，結果相當模糊的，譬如說教育程度的提昇，可能是父母花了許多時間和小孩相處，這跟學校老師做了什麼事情，一點關係都沒有。再來，各機構產出的資訊是有偏差的，也就是他們的報告都是自己想報告出來的資料以及還有許多指標，其實這些都不是由機構來控制的，再者，監督者有可能是立法機構、其它機構或是媒體，他們其實並沒有判斷的能力，由於政客並不想要有太多負面的新聞出現，負面新聞對他們是不好的，所以他們也不敢問太多問題。再來，若是用品質不好的資料來判斷，這樣是非常危險的賭注，並且是愚蠢的行爲。MCDavid 是一個談論謹慎判斷文化的作者，他會談論說我們要知道這個是資料是什麼或者它是好的資料嗎？再來，蒐集了一堆資料，但是並沒有使用，在抽屜裡面或電腦裡面都是滿滿的資料，當你在評估或修正績效時，並沒有運用到這些資料，總而言之，就是蒐集的滿滿的資料，但很少實際運用出來，若眞要談到有運用到的地方，就是所謂保護的功效而已。

現在我們來談談如何將績效與預算結合在一起，或兩者應該連在一起嗎？是否應依據績效評估的結果來判斷預算該給多少？其實，這兩者之間的關聯是非常微弱的，不是你投入的資源愈多就代表你的績效愈好，而且它可能需要花很長的時間才會顯現出你要的效果，所以你不可以只用 1 年或 6 個月的時間來評斷你的預算應該如何分配，如果說你所要達成的事情是需要花 10 年或 20 年的時間的話，這個需要的時間是很長的。再來，若將預算及績效掛勾的話，會容易產生負面的結果，甚至是荒唐的結果。如有些機構之所以可以得到很多資源是因爲一般大眾支持這個機構，這可能和它的績效沒有關係。再來，提到前面所談的公私部門的比較，在私部門中大家都不會質疑績效指標，因爲大家都覺得這個非常適用，但公部門只要談到績效指標就會有許多的爭議，譬如說資料該由誰來蒐集，爲什麼蒐集這個不蒐集那個，爲什麼選這個禮拜不選那個禮拜？因爲公部門就是一整個官僚體系，所以這裡面會有更多複雜的問題。

接下來，我們再來談透明度的問題，因爲我們將績效與預算綁在一起就無法提高透明度，所以做這樣子的連結是很有問題，而效果也是很難預期的，我個人認爲由於這兩件事的邏輯不同，週期也不同，所以這兩者無法混爲一談，在它們週期不同中，應使用不同的邏輯和不同的決策，預算通常是 1 年 1 次或 6 個月 1 次，但績效可能是更長或更短的時間，這個過程中所牽涉的人是不同的，其實編預算的人希望經費能發揮其價值，他不見得懂這個績效到底在做什麼。再者若將機構的預算和績效綁在一起的話，我們都清楚每個機構都不希望預算被刪減，所以將預算和績效綁在一起的話，每個機構都不會誠實報告他們執行績效實際的情況，他們不希望誠實報告而導致他們的預算被刪減，除此之外，也有很多機構會說我們對於這個結果的貢獻到底是什麼？其實很難評估的，譬如說用績

效衡量我們的 GDP 是沒有意義的，GDP 是受到上千個因素所影響的，所以你無法將列為單一個機構的績效指標，所以績效只是你編預算時考慮的指標之一而已，當然除此之外還有 10 幾 20 個其它因素，它預算可能你要考量的是過去歷史傳統或所謂象徵性的意義，譬如當一個新的國家或共和國成立時，它會設立一些機構，但它做的事情不一定和公務人員有關。接下來談到一些負面或荒唐的影響，假設把預算和績效綁在一起的話，會造成許多的負面效果，首先是假設你的績效時好時壞，那麼你的預算也要時高時低嗎？當它表現好時，你是要幫他加預算或減預算呢？若是減預算就是送出一個很奇怪的訊息，他們表現很好，結果預算被砍了，若機構因為某些因素而導致績效不好，那你是要懲罰它呢？或是幫助它呢？給它更多資源，若你是幫助它，給它更多資源，是否也傳出一個錯誤訊息，就是我表現不好時，反而會有更多資源。再來是，紐約的警局，他們因為轄區的不同而有不同的競爭單位，他們也是在談什麼叫做表現好什麼叫做表現不好，再來就是我們要用什麼樣的指標或標準來衡量，除此之外，另一個問題是實際在做事的人，不是提供資訊的人，雖然我提了很多的問題，但要如何改善呢？我覺得我們應要採取一些作法來減少大家的爭論，我們希望我們可以像私部門那樣，可以一起來談談要蒐集那資料？這些資料要如何蒐集？它是否反應了我們想要追求的目標？而且我們要去問實際在操作的人，我們必須清楚地知道他們要達成怎樣的目標？以及所遭遇到的困難，我們要很誠實的衡量，通常實際在操作的人最知道這個成果應該要怎樣衡量，所以你要把他們帶進來，這樣才能反應出事實的指標。第三點，我們應該要逐步建立執行的時程表，譬如說，我們有 5 年的期程，那這 5 年我們應該要怎麼走，而非追求立竿見影的成果，這樣機構才會覺得有人來支持我們，但有時候也需要顧慮到政治人物，以確保政治人物不會亂說話或者是亂移動這個目標；其次，我們也應該從作中學習，實驗哪個方式最為妥適，還有看看這些機構如何發展他們各項的能力，這部份我們很少注意，但在公職體系裡是最重要的，復次，不要用過於狹隘的報表資訊，而是要加入環境脈絡之背景因素，如我們現在要解決的問題是什麼，像學童沒有吃到營養午餐或市中心的犯罪或貧窮問題等；再者，亦應思考要用何種方式來激勵公僕，且選擇最重要的三個關鍵指標作為評量標準即可，但絕對不是 GDP，因為挑超過三個的話，絕對不是最重要的，如果你只挑三個大家會清楚知道重點在那，若你挑了 300 個，大家就不知道重點在那了，所以只需挑三個關鍵指標來回報就好了，這部分可以用新加坡為學習對象，如果新加坡的文官首長被提出嚴重控訴的話，這個文官首長是會被革職的；此外，鼓勵大眾參與績效評估 (像是智庫)，可以邀請社會大眾一起來參與監督的過程，讓社會大眾也來參與這些績效的評估，畢竟現在接受服務的對象是社會大眾，但我們卻未能將它納入，所以我們應該慎思之。

八、與談人一： Peter Vandenbruaece 處長

大家好，我是比利時聯邦公職訓練所人力資源處處長，我自己本身也是公務人員，我們單位訓練的就是比利時的公務人員，以自身經驗來作分享，也是看完 Prof. Wanna 論文之後的感想。我並沒有問題的答案，但是我們可以共同拋出一些問題，這就是找到答案的第一步了。我想談三個問題，1. 我們運用哪些人力資源發展績效指標？ 2. 有沒有可能哪些績效指標是扭曲的？或者負面的指標？ 3. 有沒有其它還可以發展哪些績效指標？我其實很同意有一些講者的發言，我們到底在訓練中要用那一些指標？我們大部份會用的都是比較量化及標準的績效指標來評估，譬如說每一年訓練多少人？訓練多少天？以及所提供那些訓練的課程，通常是用這些資料把表格填滿，這是我們在比利時公職訓練所的投入，這是很多訓練單位用來計算的一個標準，譬如說它會計算訓練經費佔了公務人員的薪水的百分比，那我們也可以看到目前為止我們經費的大約增加一倍，而經費的倍增就會看到我們

公務人力的表現比以前好，可能是效率倍增或績效倍增且服務的品質也倍增，除此之外我們也會用質化的方式來評估我們的績效評估，就像 Matthew 就有提到 4 個不同績效評估的方法，我們最常用且容易得到的資料或數據來評估，譬如說受試者的滿意度評估或者是利用測驗或考試去了解受試者的人數是否有增加，是否有可能使用到一些錯誤或扭曲的績效指標呢？有可能如此。當我回到訓練所時，我可以跟大家保證我可以增加約 50% 的訓練課程，我們不需要多餘的經費或資源，只需要將 4 天課程縮短為 2 天課程，這樣我就可以增加我的訓練課程了。訓練經費或數據所代表的是聘請顧問的費用增加了，但這或許是換湯不換藥，而現在做的事情和以前相比並無不同，但我們委外的單位費用提高或選擇了費用比較貴的單位合作，因為如此，造成訓練經費增加但結果對我們的影響卻並不一定。大部份我們在做績效評估的時候，並沒有所謂的指標去看待我們訓練的結果，再者，除了訓練之外，學習也沒有去評估，這是比較傳統的想法，因為我覺得訓練好像是把知識灌輸給人們，另有許多人傳統的想法是學生坐在座位上聽講就是學習，但現在觀點改變了，我們覺得學習的方式應該要有所改變，我過去幾天非常感謝國家文官學院的邀請，讓我們一同進行世界咖啡館這種非正式的方式學習來思考一些指標上的應用。第三個問題是有沒有其它的衡量方法呢？我覺得最大的挑戰就是找出評量重點的績效指標，那到底什麼是評量重點呢？我想跟大家分享一套機制，當我們在評量時，必須了解學員與組織的需求和問題是什麼，以及面臨的問題要如何反應在日常工作？思考訓練與職能的關聯是什麼？最後，必須思考訓練是否真的有必要？也就是我們制定績效指標的方法會大幅的改變，若依照剛才的作法及形式的話，在過去幾天的學習中，我拜會了華碩的人資長，他提到重點不是有多少人上課，而是我們能不能適時的介入提供適當的訓練內容，我認為，這是必須思考的重點。所以，我們知道績效指標的重點在於學習、人才、結果、創新及 E 化學習，所以現在使用的績效指標，它是否真的可以衡量以上的幾種能力呢？這是我們要去思考的，這張照片是我女兒和她同學，而她們的學校也會用績效指標，這就是剛才 Wanna 教授所提到，他們會用老師及學生的數量是否增加，以及他們考試的結果，還有學生對學校教育的滿意程度以及與他校比較的結果，聽到這些績效指標，有件事令我感到害怕，就是現在企業中的績效指標或在公部門的績效指標和一般學校所採用的績效指標是一樣的。這是我女兒，幾個月前她要求我帶她去買吉他，我就說 cindy，等一下，妳現在已經上了 6 年的小提琴課，而且你也上了 5 年的舞蹈課，妳沒有時間再去上其它的音樂課，你真的要買吉他嗎？她說：我想要買。我想說既然是花她自己的錢，就帶她買吧。在我出發來臺灣的前一晚，我女兒叫我到她房裡聽她彈吉他，她自彈自唱一首美妙的歌曲。若我用傳統績效評估我女兒的表現的話是無法得到適當的結果的，譬如說我無法衡量我女兒上的正式課程是零，時數也是零，她並沒有任何吉他相關的認證，但我卻可以看到她學吉他的成果，雖然她沒有正式的訓練，但她的表現還是很好。如果我用傳統的績效評估來衡量是無法得到適當結果的，我想我女兒可能常上網或和朋友出去時，就可以從朋友或同學的身上學習，如果說用傳統績效指標去衡量並非適當，而是應用最適當的方式去評量。

這張照片大家都知道這是我們會議的照片，我們已經花了 2 天的時間，我們瞭解時間的寶貴，時間也是金錢，那我們來到這裡都有它的意義，我們要用什麼樣的績效指標來評估這 2 天的會議成果？還有我們要思考為什麼要來參加這個會議？如果要用績效指標來評量我們接受的訓練的話，是要計算參與會議的數目嗎？或是要衡量我們上課程的數目嗎？或許對某一些人來說，會用量化的方式。在比利時對於參與這樣課程的參與者，或許並沒有在聽而是在上網或在工作，若在午餐過後的場次，他可能就睡著了，那是因為在績效指標當中，只是去計算參與會議的次數，並沒評估會議的結果，

所以我的問題就是我們是否應開發一些指標來反應未來要達到的目標，思考現實與理想的反應方式，我覺得我沒有辦法回答這些問題，可是我覺得現在拋出了這個問題，就像拋磚引玉一樣，希望能夠得到解決的答案，我要再次感謝主辦單位的邀請，也要謝謝大家的聆聽，我要特別感謝口譯員，因為他們翻譯的非常好，如果沒有他們的話，我們就無法有這麼好的學習，所以我們也要謝謝口譯人員，謝謝大家。

九、與談人二：張四明教授

各位貴賓，我是臺北大學的張四明，非常榮幸能夠參加本次會議。我讀了 Wanna 教授的文章，這篇文章很有意思，而這篇論文講的幾個主題就強調了我們的某些假設，也講到績效評估的執行與管理。我想要做個簡單的總結。Wanna 教授講到績效評估的理論，也就是績效評估採取的原因有好幾個，而且可能在不同的過程當中執行，除此之外並假定是一個良好的方式，而且它是良善治理的要素之一，此為大家所同意的；它也是所謂的管理核心之責任，如果沒有所謂的評估我們就無法管理，也就是說我們必須要知道我們的表現、我們身在何處，應何去何從。但現在績效管理的執行與理想有落差，就如同剛才 Wanna 教授講到指標之中有很多限制，它似乎只是一種代理的工具，無法提供給我們很有效的資訊，也無法比較各個機構的表現，譬如說我無法將橘子和蘋果相比。很多時候績效指標或績效評估只是淪為官僚的作為罷了，臺灣也有類似情形。此外，很多人對於績效指標抱持著質疑的態度，還有我們評估的資料使用也很有限，再著前述亦有提到績效指標往往淪為政府的保護機制。而我認為績效的通報也是矛盾的，因為沒有評量就沒有好的管理，所以我們需要好的指標，可是我們常去評量的東西就是我們想看的內容，此時我們根本無法達到真正的目標，因此我們的挑戰在於我們應該如何解決這個問題，我們要怎麼樣以好的方式進行績效評估，畢竟績效評估還是一件很重要的事情，只是我們面臨到很多問題。

Wanna 教授也提及績效評估與預算間之關聯太過薄弱，我也認同之。Allen Shieck 也說績效預算根本是空口白話，而如果我們想要將兩者結合一起，似乎也做得不好。我認為要結合績效與預算必須以多元的方法為之，我們不能只侷限於預算的分配、績效的獎勵或懲罰，我們或許可以將資料收集進而改變我們執行績效評估、編列預算的方式。績效評估和預算編列未必有最直接的關聯性，只是如果我們能把績效評估相關資料納入編列預算的過程中，會是一件不錯的事情。

那麼績效評估跟管理未來要何去何從呢？我覺得 Wanna 教授提出了很多問題，但並未提到太多的解決方案，這是我們所共同面對的問題。太過盲從相信績效評估的可靠性及其優點，這我也認同，我覺得我們要有互相合作的環境才能加強機關的表現，我很同意上述幾點，但現實情況卻未盡相同。我們講到互相合作環境時，例如立法委員跟行政單位間是很敵對的，所以在臺灣要有合作的環境是很不容易的一件事情。我有下述幾點建議：我認為績效管理是管理工具卻不是管理目的，它會幫我們達到目標，但我認為它就只是一個工具而已；另外我覺得透過績效評估來進行獎勵是很重要的，但我們不應過度強調物質，或是個人的目標，如獎勵或懲罰，我覺得重點不在於獎勵而是在於自我實現，例如在工作上面能夠表現得更好、獲得其他人或民眾的尊重，這是績效評估中很值得重視面向。績效評估是有很多利害關係人的，我們常講到要針對誰進行績效管理，我覺得績效評估一開始應先針對公務單位內部的人員，要看他們在績效評估的表現方面如何。最後我也同意從做中學的方法，才能夠面對挑戰，同時也可以把臺灣、澳洲和美國的經驗互相比較和交流，這是我們在此最重要的

目標所在。謝謝大家。

十、與談人三：史美強教授

　　各位貴賓大家好，根據剛剛 Wanna 和 Peter 的簡報看來績效管理的未來似乎是黯淡的，但如果細看 Wanna 的論述，其實主要挑戰的是預算分配與績效管理掛勾的問題，但我相信 Wanna 教授並非主張績效評估和目標管理分開而論。再者，Peter 分享比利時的經驗，我覺得效果非常的好，像我在小孩身上也投資了許多教育經費，似乎很難評估這些教育經費的投入，會有什麼結果。事實上，我完全同意 Wanna 教授的論點，我相信在座的各位都了解績效管理的成功之處何在，對於這個體系都是憂喜參半的感覺，那我的問題就是，我們需要績效管理嗎？我想答案是顯然的。我在這就針對 Wanna 教授的論點提出幾項回應，也提出我自己的見解。Wanna 教授提出很好的論點，他說到他並非反對績效管理，但是他擔心績效管理與報告是有偏頗的，而績效評估並未提供有效的資訊，此外，他也提到績效管理是控制管理的工具，它是控制執行的過程，如果事實如此，恐將帶來負面的後果，例如為了管理而管理，而主要目標就被模糊了。其實這些問題大家也都早已知悉，問題是我們下一步該怎麼做呢？是否可以完全排除績效管理所面臨的問題呢？或是我們完全擱置績效管理，那就像我談到的我們是不能擱置它的。我認為績效管理最後應該是與預算資源分配慢慢連結，而在那之前必須要跟目標連結起來，最後才讓績效管理和預算分配連結起來。首先要釐清績效管理的目標為何，再一步步地執行。績效管理它是一個持續要進行的過程，不該將其當成一種萬靈丹，否則將買櫝還珠，本末倒置，因此我們必須很謹慎，我們尚未達成最重要的目標，所以必須將績效管理與目標的達成建立強烈的連結。

　　以下我將提出五個建議：首先，我們應該要檢視公務體系中是否有績效管理的 DNA，我覺得績效管理這個文化其實並不存在臺灣的公務體系當中，大家只是在做表面功夫，因此我們應該要透過教育、訓練和提供誘因，來加強績效管理的文化，我們應該要獎勵公務人員達成他們的目標；第二，我們必須要不斷地檢視質化和量化的績效指標；第三，我們應該要有多種類的績效評估方法，這些方法也需要接受定期的評估、設計；第四，誰來評量績效評估的表現？我認為公務人員和大眾的投入都很重要，這對於績效管理體系的成功與否很有關係。最後一點就是透明度，這對於公眾報告績效的成果是很重要的，要讓公眾看到績效管理的成果，其實我是完全同意 Wanna 教授的論點，其中一個論點對於臺灣有不凡的意義，不過我這邊的結論是，績效管理只是工具而非目的，這個工具成功與否是看工具如何使用，謝謝大家。

十一、綜合討論：

（一）　針對前述與談回應

1. 主持人：在問題討論之前我們先讓 Wanna 教授回應前述的內容。

2. John Wanna 教授：謝謝剛剛的意見，我覺得這些意見都非常好，我也特別同意是工具而非目的本身。再來就是落差的問題，我們現在走到哪裡和我們想要走到哪裡，這當中是有落差的，這個體認是很重要的，假設我們現在在這裡，那我們想要去另外一個目標，我們不知道這個落差該如何跨過去，但是在私部門就比較容易、也比較有彈性，比如說這個主管表現差強人意，他們可以從外部去雇用新的主管，但是公私部彈性比較少，所以我們就需要有新的文化、採取新的做法。在私部門行的通的文化，如欲搬到公部門來不見得行的通。所以我們必須要建立起公部門這套能力，我們

不能越做越狹隘，這個就是我剛剛說到矛盾的地方，也就是說你所收集到的資料往往都是毫不相關的，我們希望看的更大更遠，我們要怎麼樣做才能改變現狀。有些歐美國家現在要來提倡評估服務，政府機關有什麼作法可改變來服務弱勢族群；再來我們也能與社會大眾有不同的互動過程，有些事情是可以靠社會大眾自己去做的，誠如教育，也就是鼓勵大眾在教育裡多一些參與和付出，例如說社會有些問題看似是無法克服的障礙，但是我們可以一起腦力激盪，思索我們或許能給人民不一樣的力量，也許有些事情不是政府可以做的，反而是民眾可以做得更好，像是新加坡的文化在這部分就做得很成功，而西方自由主義國家在這方面就做得不是那麼理想，因為大家都覺得這是別人的問題，這應該由別人來做，就像今天早上講者有提到 James town 以及普利茅斯的問題，一個獨立派另一個尊重社會群體，兩種不同的思想，兩種不同的派別。

（二）提問

1. 提問人：Massimo de Cristofaro、義大利公共行政學院、資深研究員

提問要旨：在我們的公共管理國家行政學院會衡量訓練的效果，但我們卻未能妥善關注機構裡反彈的力量。所以我們要問的是，這些文官、經理人是真的全心投入要去做改革，因為如果沒有他們的參與及投入的話，就如臺灣的總統說我們有很多改進的空間，而且他也很重視公務人力資源的發展，但除此之外其他有多少人想改進呢？

2. 提問人：朱玉嬋、教育部、國際文教處副處長

提問要旨：我想聽聽 Wanna 對大學全球排名的意見，有 4 到 5 項指標，這些指標很大幅度地影響某些國家的做事方式，有些國家甚至還會使用國家資源來幫助某些大學進入全球大學評比的排名，我想要請問您對於世界大學評比未來的發展為如何？

3. 提問人：吳榮泉、外交部、大使

提問要旨：我在過去 10 到 15 年必須要做我們單位的績效評估，我覺得我們評估的指標沒有什麼不同，所以，我會去進行我同事的評分，進行他們在忠誠度或能力方面的評比，意指從過去到現在我們的績效指標並沒有太多的改變，我覺得我們似乎只是在評分而已，並不重視所謂的績效指標的重點為何，我想要問 Peter，你們的訓練所如何去收訓練的學員，是由他們推薦或是各單位自行推薦呢？，以及你們訓練的學員後來有多少人得以升遷，又這些升遷與訓練之間有任何關聯嗎？

（三）回應

1. 回應人：演講人 John Wanna 教授

回應內容：

(1) 我覺得你指出了非常重要的問題，我們該怎麼樣去衡量訓練的結果和影響。當然在公部門裡面要維持現狀守成是很習以為常的一種文化，很難讓人用別的方式來做事，那麼在公部門的主管之所以能爬到如此高階的位子，其實就是靠秉持保守與避免風險的態度，所以他們的文化裡就是避免冒險，但也有所謂變革型的領導者，我們認為公部門還是會改變的，公私部門和第三部門（如慈善機構）現在的界定都劃分得很清楚，但我認為未來的公部門共事對象將會更加多元化，例如小學老師共事的對象還包含志工，這是一種混合的模式，未來在公共服務的領域裡面，這種現象可能會愈趨普遍，也就是說未來公共服務的提供將會是更為混合型的，政府都在追求所謂的網絡，亦即大家必須要知道該如何管理這些資源網絡、人際網絡，

單就這點我們有很多的著墨空間。

(2) 我覺得評比會同時影響政府、父母和學生，如果我們比較各　國大學是沒有意義的，因為我覺得評比的重點已經不在大學上面了，畢竟大學是很鬆散的單位，我的學校在世界排名其實也不低，當然排名越強越好，但這個大學評比將以兩種方式影響政府，其一可能是政府會給予表現好的大學較多補助或資源，我覺得我們應進行領域的評比而非大學的評比，現在在美國就是在評比大學的領域，所以大家現在到國外去唸大學的時候，是唸一個比較強的領域，而不是評比比較好的大學，畢竟蘋果和橘子是很難比較的。

(3) 在此我必須要再補充一點，我覺得你所提出的問題跟我的觀點很相近，評比指標其實是很無用的，我覺得我們不應該只是去填表格而已，我們應該要與人員之間進行面談，我們剛才講到公私部門在做績效評比的時候，必須要去了解單位當中需要的資源是甚麼，對人員亦須進行面談等等；我們要去詢問未來五到十年的目標是甚麼；此外我們或許可以讓教授休假，讓他們可以去跟第三部門的人互相合作等；讓公務人員去海外生活，如此才能更正視自身文化並適應環境；我們也希望社工人員能夠有 1 至 2 年的時間去國外的非政府組織工作。我們希望能更為宏觀的角度去看待公務人員未來的發展，而這與您所提出的問題是相關的。

2. 回應人：與談人 Peter Vandenbruaene 處長

回應內容：

(1) 讓我再說明比利時的例子，我們要怎麼樣保證高階文官願意把結果納入考量，就是我們提供的標準訓練是屬於一系列訓練當中的一環，然後我們也有客製化的訓練和課程。而我之前提到的訓練需求分析，這是我們和政府人事發展部門的共識，就是我們要先找出訓練想要達成的目標，並了解接受訓練的利害關係人到底是誰，也就是說，某一個機關在訓練之後該如何確保因訓練而帶來的轉變，但如相關單位無法配合的話，我們就不會進行這個客製化的訓練，因為我們認為我們過去接受太多的訓練，而且花了太多的經費，去打造客製化的訓練課程，但是我們不知道實際的利害關係人是誰，或是那個組織有沒有人願意去負責，所以後來這個制度就改變了。

(2) 我先補充一點，我覺得您提的論點就是和我的觀念很像，就是有很多的評比指標它是很無用的，我們不應該只用表格來評比，而是要與相關人員面談，如我們所談公私部門間績效評比的時候，有談到我們必須要了解單位的需求及相關人員進行面談等等，再來是 5 到 10 年的目標是什麼，我們是不是需要訓練課程，此外，我們是否可以讓教授休假讓他們可以和非政府組織進行合作等，也就利用合作的方式，就像外交人員他必須要進行更多的合作，昨天有談到，許多公務人員他們應該到海外去體驗，這樣可以讓他們更正視自己的文化，適應新環境，之後，再加以升遷，那我們也希望我們的社工人員有 1-2 年的時間可以到不同的慈善機構去工作，這點當然不容易，因為很多公務人員他們面臨許多的問題，比如說到那裡上班會有交通不便的問題，提出許多鎖碎的理由等等，但我們希望能利用較宏觀的觀點來看待他們的職業發展，這跟您剛所提的滿蠻相似的。還有，我們的學生在尚未畢業前就先找到很好的發展，很多學生的發展不一定只有公職，他們也往學術方面發展。我盡量簡要的回覆您的問題，您所提的問題相對複雜，首先就第一個問題，其實我們不會去篩接受訓練的學員，學員參與訓練的資格其實是由部會當中特定學員跟他們的主管討論的。除此之外，透過績效的評估，公務人員自己亦得自發性地參與訓練課程，可是在這個之後，他們就會到中央的人事機

構，如果中央認為某些學員是需要接受訓練的，訓練可以幫助他們達成職務上的目標的話，他們就可以參加我們的課程，我們的訓練所就會邀請他們訓練，所以我們自己本身並不會篩選學員，我們覺得我們並不能作篩選，這個應該是由各個單位所決定的。第二個問題問到我們的學員晉升的比例，或是接受訓練與晉升的關係，其實在我們的體制下晉升有兩種方法：一、自動法，也就是說只要服務滿六年薪資就會自動往上，不過七年前這個制度已改變。而我們的訓練當中，我們有認證訓練課程，學員在這個課程結束之後必須接受考試，如果通過考試，只要有六年經歷即可晉升，如果未能通過考試，他就沒有辦法晉升，但他可以等待一年時間，隔年再考；第二種晉升方式是透過政府單位較高的職缺公布給符合資歷的公務人員來參加，符合資格者再由委員會來決定，這是比較快速的晉升方法。

十二、主持人結語

　　績效管理很重要卻又不易達成，但如規劃得當是能增加政府效率與效度的。

PERFORMANCE REPORTING &
PERFORMANCE MANAGEMENT –
Challenging the Logic of Linking Performance Regimes to
Budgetary Decision-making & Resource Allocation

John Wanna
ANZSOG, ANU, Australia

What are our Challenges..?

- Want to link strategy to investments in resources, & effectiveness (value for money)
- Notion that Performance Management (PM) is the bridge
- PM is 'holy grail' for govt, an ideal, quest
- Performance appraisal accepted as a 'good'
- Beyond question or dispute, only Neanderthals oppose…
- But we should question these claims…

Current Concerns at How Perf Man being Applied & Imposed

- Underlying <u>assumptions</u> – too much faith in benefits, assume chosen indicators provide valuable information
- <u>Implementation</u> of Performance Monitoring is flawed, poor indicators, distortionary, perverse, conspicuous, external imposed, central & standard
- <u>Outcomes</u> of PM – poorer evaluation techniques, not got good review capacities, limited use made, but mirage of performance budgeting alluring

the Australia and New Zealand
School of Government

Reasons for PM – seem clear but…?

- Normally for improvement of 'x'
- Rationale for policy interventions
- Good implementation planning
- Monitoring for continuous improvement
- Establish competitive ranking scale
- Justification and defence (protection on PS)

Different reasons lead different PM designs and many reasons can be mixed together

the Australia and New Zealand
School of Government

Performance Management in the hands of others...

- Central bureaucratic exercises, self-referencing, remote
- Some govts announced central plans to evaluate all programs (? impractical)
- Power, controlling the reins (emp/disemp)
- Conducted for different actors & purposes
  - Executive, legislative, public, media, professions
- Influenced by different waves, fashions, methodologies

School of Government
the Australia and New Zealand

Performance Indicators

- Often surrogates, visible, quantifiable
- Govt ask: what can we get, what we have...
- Neglect causal links Indicators v Outcomes
- PI = levels of activity, 'busy-ness' metrics
- or PI = process data, volumes, queues +/-
- or PI = targets, ratios, costs/benefit
- Capture some quantities, harder on quality
- Mostly measure 'effort' not achievement (paradox)
- Seduction of the 'table', anything in words can't be tabularised, numbers sought but mislead

School of Government
the Australia and New Zealand

Myths Associated with Performance Management

- Trust in govt will increase?

- That the public will care, or govts will care?

- Creates more innovation?

- Leads to more efficiency?

- Rewards (Ind/Org) improves performance?

- That PM motivates employees/providers

- Policy is evidence-based & initial objectives clear & so amenable to statistical review?

the Australia and New Zealand
School of Government

Problems Judging Performance

- No clear causal link, PI to outcomes opaque
- Much data produced by agencies (bias, innocuous)
- Many indicators not controlled by agencies
- Have overseers requisite judging skills, subjective
- Making judgments on poor data is risky
- Do govts actually use performance data when reviewing, re-authorising, amending
- Very limited use is made, mostly cosmetic, hence comment it is 'protective device'

the Australia and New Zealand
School of Government

Problems linking Performance to Budgets – 'Performance Budgeting'

- Very weak links exist
- Long lead times required
- Long reporting lines
- Often perverse consequences
- Results Perf not tied to resource allocation
- Arguments about measurements, results
- Can we demonstrate transparency

the Australia and New Zealand
School of Government

Linking Performance and Budget – why not to do it...

- Linking is problematic and unpredictable
- Likely to see *neither* achieved well
- Different logics & cycles & players
- Creates incentives to disrupt or game-play
- Impossible to measure an agency's sole contribution to an outcome
- Performance is only one factor in budgetary issues
- Budgets=rationing, legacies, spending, symbolic

the Australia and New Zealand
School of Government

Perverse Consequences

- <u>Should</u> we reward performance and how?
- Ind/group/agency? Unintended consequences
- What <u>if</u> resources linked to performance?
- If agency over-performs – cut or reward?
- If agency underperforms – censor or assist?
- What criteria counted? Lead to focus on things <u>counted</u> at expense of others
- Agent not principal generates information – leads to asymmetry and gaming

the Australia and New Zealand
School of Government

So How Better...

- Minimise scope for arguments
- Ask deliverers, ask those closest how measure, know
- Build in agreed imp schedules for PM improvement
- Learning by doing, experimental, how are agencies cultivating their capacities, anticipating futures
- Don't use narrow output reporting, frame narratives, mapping, client journeys
- Think of ways to motivate PS – celebrations, off-line
- Pick 3 most important high-level measures
- Use hard complaints systems with sanctions, CEO
- Engage public, outreach agencies, constructive reviewers, ask for community assessments

the Australia and New Zealand
School of Government

327

Comments and Discussion

School of Government
the Australia and New Zealand

Session: Utilisation and Limitations of Performance Indicators

Performance Reporting and Performance Management – Challenging the Logic of Linking Performance Regimes to Budgetary Decision-making and Resource Allocation

John Wanna

ANZSOG, ANU, Canberra Australia

Arguments

The main arguments of this paper concern: (i) the assumptions underlying performance monitoring in public sector environments, (ii) the implementation of performance measurement in practice; and (iii) the outcomes of performance management in government.

In relation to the underlying assumptions, I argue we invest too much blind faith in the benefits and reliability of performance monitoring; and that we assume and often take for granted that certain selected indicators can provide valuable assessments of the performance of public delivery actors.

In relation to implementation, I suggest that we do not yet have a good grip on measuring and assessing performance in the public sector. We have developed relatively poor indicators and measuring metrics and many of the performance indicators we have are often distortionary (and possibly perverse). Indicators also tend to capture the conspicuous phenomena and observable criteria not necessarily the substantive issues or measures of effectiveness. As a general rule, performance measuring regimes tend to be self-calibrating rather than tailored to the specific needs of programs or their context/starting points. Moreover, those responsible for establishing performance monitoring regimes in the public sector still tend to prefer centralise systems of monitoring, seeking to standardise measuring requirements and reporting formats, and establishing generalizable monitoring systems when an allowance for greater diversity may better capture real levels of effectiveness.

In terms of the outcomes of performance management, I would argue we tend to have even poorer evaluation and review capacities when judging performance–related data. When we produce performance information it is generally the case that limited use is actually made of such information by governments and relevant overseers (and almost none by the citizenry). And, when thinking of future developments we should be very wary about trying to link performance monitoring and reporting of activities too closely with subsequent resource allocation, budgetary decision-making and the alluring mirage of 'performance budgeting'.

Reasons for Measuring Performance

Performance in public sector agencies and programs is not measured for its own sake. Rather assessments of performance are undertaken with some pious hope of improving certain policies, programs, activities or operations, organisational achievements and community outcomes. Performance monitoring and reporting can be undertaken for various reasons and at different stages of policy processes: it can be a rationale for policy interventions (initial motivation); it can be a deliberate aspect of good planning and design (formulation); it can be a form of comparative assessment (imposing competitive pressures or rankings); it can be a device for urging continuous improvement (productivity or effectiveness), or a defence and justification (a shield). However, in practice many examples of performance monitoring are patched-up exercises, very ad hoc and partial in nature and imposed after the event. Policy deliverers may have external indicators imposed upon them, sometimes relying on metrics or processes on which they were not particularly focused.

Yet, performance assessment (like coordination) is thus largely an assumed ‘good’, necessary beyond question or dispute (although there may be vehement disputes over the types and methods of information gathering). Many public management texts routinely present performance measurement as a vital component of good governance, an essential part of defensible policy processes, and a core responsibility of management. It is also seen as an important dimension of public accountability for service delivery and the deployment of resources, authority, equity etc.

However, it should be noted that there is a strong tendency to centralise and standardise monitoring systems, to attempt to ensure comparability in reporting formats. But comparability can be a straitjacket. Performance measuring regimes in the public sector tend to be run as bureaucratic exercises, becoming self-referential and self-calibrating rather than premised on the specific needs of programs or their context/starting points. Performance regimes are thus expressions of power relations, potentially empowering some actors while disempowering others. Political executives and legislatures can impose performance regimes to ‘control the reins’, impose their particular priorities or extract savings. Executives may invest in performance monitoring as a way of protecting themselves and managing political risk. There can be a degree of distrust and cynicism shown towards performance measurement.

Performance reports are potentially utilised by different actors in the policy process. Ministers and politicians may view them as evidence of policy effectiveness or the net worth of additional resource investments. Some managerial actors may use information for executive feedback on the progress and performance of policies/programs. It is not necessarily a public process or transparent. Accountable actors (such as legislators, scrutiny bodies or auditors-general) may view performance data to gauge efficiencies and effectiveness to the extent they can. The media can use performance information in producing newsworthy stories, although only ‘bad news’ items tend to appear

regularly. Unions or professional associations may use performance data to help regulate their employment practices, workloads or training requirements.

There have also been different waves of interest internationally in performance measurement. Many nineteenth century western governments were motivated by parsimony and imposed economies. Americans became interested in organisational performance ahead of most jurisdictions with the 'performance movement' of the early twentieth century feeding into broader reform efforts. Management was motivated by PPBS, MBO, ZBB, and later by performance based laws (GPRA, PART). This was accompanied by the development of extensive cultures of evaluation in post-war public policies – especially through randomised sampling and the highly quantitative use of management-derived surveys (open to agency bias) (see Radin 2006). Many other advanced nations began to focus on efficiency and technological improvements from the 1960s, with many developing nations following suit from the 1990s.

When governments embraced business-like practices from the 1980s onwards (managerialism, New Public Management, devolution, outsourcing) they often embraced performance measurement as a co-requisite to indicate improvement in management practices and outputs. Many governments announced extensive plans to evaluate programs (the holy grail of policy evaluation and program evaluation, which has often proved elusive, see Wanna et al. 2010). Over time, there has been a steady improvement in reporting arrangements and in their consistency, and generally many delivery areas of public policy now have pre-specified and timely reporting requirements; however such progress has not been linear (there has been some backsliding and gaming – especially between playing off input reporting versus output reporting, and there remain questions over the quality of the information captured). Performance measurement even under more flexible business models of administration remains a 'work-in-progress'. In some jurisdictions arms-length monitoring bodies have also become more significant – and can name and shame agencies publicly.

More recently, many governments internationally have committed themselves to realign public services through citizen engagement to better reflect citizens' preferences and satisfactions. This is ostensibly a means of reconceptualising performance from a client-oriented perspective. However, to date there is little evidence that citizens' expectations, citizen surveys, or satisfaction indices have been widely used in seeking feedback concerning the quality of services delivered by governments. There are some examples of nations undertaking regular citizen surveys of satisfaction (eg New Zealand, Canada) but not much close integration of this information with line delivery.

How do we Develop Performance Indicators

Indicators of performance are often surrogates. We grab what we can; we collect the collectable; we measure the measurable. Indicators have many limitations. If indicators are merely surrogates they may be readily collectable but tell us very little of importance. Criteria we can count and

measure tend to be measured, and therefore manufactured by producers. There may be no causal link between the indicator(s) collected and the desired outcomes expected. Agencies may simply 'tick the box' to comply with the collection of data but it makes no difference to the performance of the organisation. Perverse logics also apply in that many agencies will be tempted to generate results only in those criteria measured which may entirely distort their behaviour and in some cases be counter-productive. For instance measuring police performance by arrest rates is a simple enough metric, but it provides incentives to police to arrest citizens for any slight misdemeanours, to record all arrests or make multiple arrests, and to show little discretion in making arrests. It can have other unintended consequences such as filling the courts and gaols with relatively non-serious offenders. Police may be tempted to arrest and imprison fine-defaulters who are not really criminals.

Hence, crude performance indicators are often restricted to easily identified and observable criteria, but may be limited in their actual applicability and usefulness. Many of such indicators only indicate levels of activity, or the 'busy-ness' of agencies/deliverers (they are volume measures). Other indicators can capture process data (such as the number of meetings held, cabinet submissions processed, inspections undertaken), which provides a measure or public service routines or throughput. Quantifiable indicators can be presented as positive or desirable achievements, or as negative features to be lessened or avoided. Positive indicators include desired targets, stretch targets or number of processed items completed; workload estimates; program outputs. Negative indicators include such measures as backlogs, queues and waiting lists, losses of clients or consumers (say declining educational enrolments). Some indicators express ratios of inputs to outputs, but as productivity measures these ratios are dependent on many other assumptions and calculations – eg, what fixed and variable costs are included in the inputs (including overheads) and what data on outputs can be generated. For instance, the costs of running a public school maybe itemisable, but what departmental overheads and central government costs should be allocated to the school? Moreover, apart from such measures as school attendance indicators, what outputs in terms of quality of education can be quantified to weigh against the consolidated input aggregates.

Additional problems with indicators are that we often find it difficult to produce or collect any reliable or relevant data – we frequently do not collect data on things governments are interested in or commit resources towards. For instance, Australian governments commit funds to improve the 'school readiness of children' yet there is no definition of what 'school readiness' means and no data collected on indicators of readiness. It is not even clear what kind of data might be collected in the first place. In some cases, governments can produce some data but it is of poor quality, unreliable, distorted, and non-comparable over time of between agencies/units. Data comparability can be impeded when actors use different definitions of concepts or designated activities (eg, prison rehabilitation) and complicated by different collection methodologies. Data can be captured at different times or across totally different contexts (socio-economic, regional, ethnic or indigenous

communities). Delays in processing or reporting data from the front line (compilations of which can sometimes take years to see the effects) can erode the data's usefulness in judging the effectiveness of policies/programs or in policy/program redesign.

Making judgments on poor quality data is inherently risky and problematic. The relationship between different sets of indicators pertaining to performance (say waiting times in hospitals versus health outcomes) are not necessarily straightforward or consistent. Improvements in one measure may come at the cost of another, or the measures may have almost nothing to do with each other. Some governments in introducing performance measuring regimes have called for far too many indicators of performance (the multiple indicator malaise) – such that agencies spend inordinate amounts of time chasing data on indicators and managers do not know what expectations they are meant to be managing towards. Indicator collection can become a veritable industry, and governments or legislatures tend to multiply the number of indicators required, piling demands for more information on top of other requests (including for accountability and corporate requirements) without culling indicators to a feasible minimum.

As a general rule, the more we collect data on discrete activities and the associated outputs of an agency the more we can produce quantifiable indicators (but subject to the above qualifications); the more we seek to collect data on outcomes and the effectiveness of programs the more problematic will be the indicators. Indeed, many governments do not collect data on outcomes they espouse as their objectives (eg, environmental sustainability, safer communities and improved community capabilities, and improved social cohesion).

There are also special problems in both federal and unitary nations with intergovernmental transfers and shared responsibilities for policy sectors. In federal nations sub-national governments will often accept funding proposals initiated by the central government without necessarily being committed to these policy objectives. Hence, considerable gaming can occur (eg, over precise commitments, matching funding, the timing of resource commitments – front-ending versus back-ending); and some jurisdictions can regard gaming is an expected and legitimate part of intergovernmental relations because sub-national jurisdictions may have limited rights to taxation and no guaranteed shares of the taxes raised from their own residents.

In nations like Australia and Canada there is limited evaluation of performance where funding crosses jurisdictional boundaries, because neither level has the formal powers or sometimes even the inclination to monitor and hold to account. Their legislatures rarely investigate (if ever) policies/programs/grants that are specified by the central executive, but implemented by state or provincial governments; nor do their principal accountability actors (parliamentary committees, auditors-general, commissions of audit; performance or cost commissions). And if constitutional arrangements prevent substantial institutional re-design, then it is hard to make significant political

headway over performance in areas of shared responsibilities. While various occasional pleas for reform can be made by actors (such as separating roles and responsibilities, rebalancing VFI, moving to guaranteed revenue shares, transferring responsibilities entirely one level of government) they have had little practical effect. Both these jurisdictions have only just enabled their respective auditors-general to investigate and audit performance over cross-jurisdictional funding arrangements.

Myths Associated with Performance Management

There are many myths associated with performance management in the public sector. These include that trust in government will increase or improve if governments perform better. This assumed association is neither a one-to-one relationship nor necessarily a positive relationship. Improved services can inflate community expectations and lead to disappointment and greater distrust. And as governments have attempted to show performance improvements in policy areas, it has often bred increased cynicism and disbelief.

Some believe that innovation creates better performance, but again the relationship is ambiguous. Innovations aimed at increasing effectiveness may come at the cost of efficiency. Innovations can increase cost-structures or timeliness – especially bottom-up innovations which may simply transfer transaction costs from providers to receivers. Moreover, major technological innovations can be costly investments, slow to implement, and take years to evaluate the benefits. And as Dunleavy has argued major ITC innovations can undermine productivity in the short to medium-term, produce bureaucratic cultures of resistance, and be difficult to evaluate using a cost-benefit analysis. Similarly, many managerialists argued and believed that increased flexibilities in public management would lead to increased performance, yet when it came to demonstrating improved performance were generally unable to convincingly prove results. For instance, the Australian parliament repeatedly asked the government to produce reports evaluating the public sector reforms of the 1980s and 1990s, and although four reports were produced over a decade they were largely silent on performance improvement.

Personal incentives to improve performance are equally problematic. Rewarding high performers or high-performing units (branches, schools) can lead to the opposite effects system-wide – through discouragement of low-performers, envy, perceptions of fairness and unfairness, disputes over criteria/possibilities. It is not clear that rewarding high-performing entities actually leads to improved performance in those entities, and fortune/happenstance may have been an important contributor to enhanced performance levels. Performance pay for officials (or bonuses) again does not often lead to heightened individual or agency performance, and in their application can produce resentment, an undue emphasis on individualism, inflated exaggerations of personal contributions, team discouragement, even distorting of data and indicators. There can also be criticisms of the

politicised system of reviewing and awarding performance bonuses, which may reward factors not related to performance (seniority, reputation, sensitivity, gender, regional-urban etc). This raises the whole issue of whether we can reward public sector people for good performance – it is not an easy question to answer.

There is also a powerful myth that modern public policies are evidence-based and supported by factual empirical assessments of circumstances, and that this can then (and should then) be tested to evaluate the value of the policy intervention. Performance measurement then is a form of policy evaluation. However, policies are still politically driven and governments have invested considerable 'face' in their policies, meaning that unflattering evaluations threaten to embarrass the government or selected ministers. Under such conditions policies may evade review or be only partially reviewed to avoid criticism of the policy intent or design.

What do we get from Performance Measurement – the Tricky Task of Judging Performance

Perhaps the most significant problem with performance indicators is that the connection between the reported indicator(s) and the desired results is generally opaque, and often the subject of much disagreement. There is often no clear causal link between sets of indicators and the objectives of the policy/program. We can readily capture reams of data but much of it may not tell us much about performance; and some may mislead. Moreover, the indicators may not be controlled by the agency or program operators; and many other factors beyond the control of the agency or program operators may contribute to the measure. Hence, as McDavid (2006) has warned, reading off judgments of performance from flawed data or from information that does not tell the overseer or reviewer what they think they are getting. He recommends developing cultures of prudence and realistic appreciations of what is possible given circumstances.

Secondly, there is the question of whether the overseers and evaluators have the requisite skills and experience to review performance. Do they have sufficient content knowledge and experience with similar policies/programs to be able to credibly evaluate similar policies/programs? Staff in specialist review bodies quickly lose their 'groundedness' and ability to empathise with implementation realities. Experienced joint-management reviews were once fashionable (consisting of elite peer reviews by independent experts) but have since tended to fall from favour (except with project management – eg, Gateway Review methodologies). Generally, the review capacities of government in judging performance–related data are not great or verifiable (independently reviewed or audited).

Then, thirdly, there is the issue of whether governments actually use performance data when making policy choices, re-authorising programs, re-committing resources to ongoing programs.

Arguably, only limited use is actually made of performance information by governments and relevant overseers when judging policy outcomes and re-considering approvals or extensions (and unfortunately there is almost no direct involvement by the citizenry in policy/program review – it remains almost an entirely perfunctory executive prerogative).

The Logic of Linking Performance Measurement to Budgeting and Further Resource Allocation

There has been much recent rhetoric associated with the need to link performance measurement with budgeting and resource allocation. This invokes the notion that the budget process can be structured to enhance performance/productivity and also demonstrate better performance in reporting. Yet, the links between budget decision-making and performance remain weak (if existent at all) and may have counter-productive or perverse consequences (eg, an increase in prison escapes contrary to performance objectives may result in political embarrassment and the allocation of additional resources to prison managers – and sceptics may consider that there may also be an incentive to be lax with security).

Many governments have moved away from line item budgeting to framing the budget around expected results (eg, outcomes, outputs, key performance indicators). Budget documentation has been extensively revised to provide results-based frameworks against which resources have been allocated. The impression given is that the government has weighed value for money considerations against its expected results and allocated resources accordingly. Cynics might argue that these governments have allocated their resources along traditional (legacy) lines and then presented them cosmetically in a result-based packaging. Certainly more performance information has been included in recent budget documentation (whereas little was included some decades ago) but it is not clear that an assessment of performance informs the basis for resource allocation.

If we see performance monitoring as a means of giving legislatures and taxpayers accountability information on the value for money of public provision then there are still major gaps in the governmental systems. Governments have talked about performance for over 30 years now and have accepted results-based strategic plans and budgetary documentation, but are reluctant to release detailed information and assessments of performance. Some jurisdictions provide more performance-related information in intended planning documents but such intentional commitments may not then be used for reporting purposes. Often what governments report is politically motivated, the parameters, indicators, or outputs can be changed by the executive to disguise lower than anticipated performance results.

These developments have encouraged some governments recently to talk of adopting 'performance budgeting'. But the concept is problematic for many reasons. Allen Schick began a study in the US entitled 'Does Performance Budgeting Perform?' and tried to disentangle the logics of linking the two processes together. He quickly found that performance budgeting was largely a hollow commitment and that there were insurmountable problems to be found in evaluating its performance as a new system of effective resource allocation.

Although there is currently a huge effort going on to link performance issues into the budget process, I would suggest that this trajectory is misplaced. Linking the two fields of budgeting and performance is likely to see neither achieved well – they essentially run on different and discrete logics. We may undermine the integrity of budgeting by insisting performance-related evidence is incorporated and used in some way as a basis of decision-making; and conversely, we may undermine the integrity of performance improvement if we try to run it through the budget process. In short, I'd argue that contrary to notions of moving to performance budgeting – the degree to which the budget is a suitable mechanism for examining performance is highly questionable.

Some of the more concrete problems associated with performance budgeting are: how might we reward or discipline parts of the public sector in a results-based budget framework. It is always possible to incorporate performance-related information into the budget process, but there are some real problems in using this information to formulate budgetary allocations. How might jurisdictions implement performance budgeting, would it make budgeting overly complicated and political, and what other problems would they encounter? Do we set minimal quotas, performance indicators or stretch targets? How can performance be accurately measured when line agencies will be largely in control of the performance information fed back into the process?

In the private sector there is relatively little argument about performance measures – but they are simpler and there is far less disagreement on the important indicators. In the public sector, the concept of performance and the nature of measurements (or findings) is highly contested and the subject of intense disputes. There is no agreed basis of comparison over time of between agencies. Also governments have to perform as well as they can in qualitative functions (eg, economic policy advice, the assessment of the needs or training requirements of the unemployed, the facilitation of a competitive environment) or in areas of society in which they have been forced to act or take on a role (eg, law enforcement, social assistance, treatment of the sick, housing the poor). If an agency over-performs do we reward it or cut its budget? If an agency underperforms do we censure it, reduce its funds, or allocate additional funds to help lift its performance. Is the answer different if a senior or indispensable minister is involved? In any of these cases the behavioural incentives lie essentially with the agent not the principal (as does the information). There are no easy answers to this conundrum – but perhaps we should try to address performance issues in their own terms not as part of a resourcing process. Some may respond that if we do not put performance issues into

the budget then any actions that follow will not have teeth, but this suggests that the budget process can be used in such a way without generating perverse consequences, and that we cannot seek to improve performance even while leaving budgetary considerations to one side.

In some jurisdictions we are seeing governments attempting to reward or sanction delivery agencies for satisfying pre-determined performance standards. Their thinking is that delivery agencies will respond to behavioural economics and achieve higher levels of performance to receive additional performance payments for their programs. There can also be sanction penalties associated with these initiatives, but already we know that the process of actually imposing sanctions is difficult and politically problematic for various reasons (imposing penalties on other jurisdictions, imposing penalties on political or administrative colleagues, imposing penalties on front-line operators etc). Many public sector experts believe that sanctions do not work and that penalty provisions are deleterious to good performance or improving performance.

Conclusion

Let me emphasise I am not against performance reporting and performance management. Rather, I am concerned that many of the systems we use are partial, distortionary, ill-equipped for purpose, not good at conveying important information on effectiveness and outcomes. This is not to suggest that governments should cease attempting to improve their performance measurement and reporting regimes, but that governments should look to the bigger picture of performance improvement, the lessons of policy learning, better policy design and improved implementation. Governments collectively need to explore what factors are driving performance results (positive or negative) and use these actively in pursuing a learning agenda. I would argue that a cooperative and no-fault environment is important to encourage real improvements in performance to materialise (rather than an adversarial, insular, defensive mindset interested in 'gotcha' politics and scapegoating). Hence, a different inter-agency culture may be required, along with greater initiative from amongst those organisations and agencies most directly involved in service design and delivery.

As a general point, I would suggest simply imposing more performance reporting requirements will have limited benefit to the overall public policy process, and it may harm actual output levels and the search for greater effectiveness. Improvements in measurement, transparency (show and tell or shame) and performance reporting can only go so far. They are relatively sterile measures, often punitive in rationale, retrospective, likely to discourage innovation or risk-taking, and likely to enhance the incentives for gaming within the system as a defensive instinct. It is also onerous to comply with the accountability burden for no apparent gain in performance or community benefit.

There are great possibilities for oversight bodies and legislatures to work with providers and delivery agencies to assess delivery chains to explore where significant improvements can be made

and where governments could make better investments of resources. But this would require the development of better relationships between these bodies, based on mutual endeavour, cooperation and a common interest in resolving problems. Many of these oversight bodies as institutions stemmed from the hostile scrutiny accountabilities coming out of the nineteenth century, highly political, adversarial and blame-oriented in character. Initiatives such as joint performance auditing, constructive joint-management reviews especially of projects, learning from doing and from bottom-up experience, all have potential to improve policy design and delivery. Legislatures could also be encouraged to take on a more managerial oversight role rather than a tribunal or inquisitorial role. There is much scope for upper houses and parliamentary committees to move in this direction and undertake some of this work as well as for other specialist policy-review bodies, think tanks, and even community review mechanisms.

References

McDavid, J. 2006. Program Evaluation and Performance Management, Sage.

Radin, B.A. 2006. Challenging the Performance Movement, George Washington Press.

Schick, A. 2005 'Does Performance Budgeting Perform', ANZSOG Presentation, Canberra.

Wanna, J., Butcher, J. and Freyens, B. 2010. Policy in Action, UNSW Press.

HRD Performance Indicators at the Training Institute of the Belgian Federal Administration

Peter Vandenbruaene

HRD performance indicators

3 questions:

- 1. Which performance indicators do we use within the field of Human Resource Development?

- 2. Is it possible that these indicators are sometimes "distorting" and even "perverse"?

- 3. Are there any alternatives?

HRD performance indicators

Question 1: Which performance indicators do we use within the field of Human Resource Development?

Quantitative:

1. number of civil servants trained per year

2. the total number of training days

3. the number of courses on offer

4. the number of hours of instruction or training individual civil servants have to get over a certain period of time

5. total federal training investment compared to that of other organizations in the public or private sector

HRD performance indicators

Federal Training investment in the Belgian federal civil service

= % of the total mass of salary of the Federal Public Services spent on training its civil servants

2002 : 1,23%	2006 : 2,04%
2003 : 1,72%	2007 : 2,69%
2004 : 2,14%	2008 : 2,28%
2005 : 2,34%	2009: 2,94%

Service public fédéral
Personnel et Organisation
Federale Overheidsdienst
Personeel en Organisatie
.be

---

total federal training investment as a whole with that of other organizations in the public or even private sector.

HRD performance indicators

Question 1: Which performance indicators do we use within the field of Human Resource Development?

Qualitative

Training evaluation data mainly based on:

1. Participant satisfaction (« happy sheets »)
2. Knowledge / skills acquired (« tests »/ « exams »)

Service public fédéral
Personnel et Organisation
Federale Overheidsdienst
Personeel en Organisatie
.be

total federal training investment as a whole with that of other organizations in the public or even private sector.

HRD performance indicators

Question 2. Is it possible that these indicators are sometimes "distorting" and even "perverse"?

- Are they an indication of what really matters?
  E.g.: Total number of courses organized: +50%, +100% ?

- What do *total training investment* percentages really tell us?

- What about transfer and impact?

- What about learning instead of training?

HRD performance indicators

Question 3. Are there any alternatives?

Performance indicators based on <u>what really matters</u>:

- Training Needs Analysis (4 level gap analysis)

- Transfer and Impact evaluation (3rd and 4th level)

- « return on investment » & « return on expectation »

- Outcome instead of output

HRD performance indicators

Question 3. Are there any alternatives?

Performance indicators based on <u>what really matters</u>:

- Training Needs Analysis (4 level gap analysis)

- Transfer and Impact evaluation (3rd and 4th level)

- « return on investment » & « return on expectation »
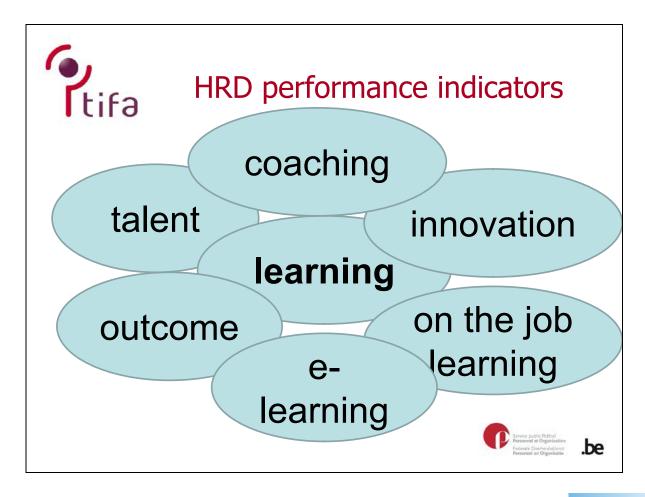
- Outcome instead of output

HRD performance indicators

Question 3. Are there any alternatives?

"it's not about the number of people taking a course,

it is all about comparing interventions that should happen at the right time, for the right people at the right place

with the real time, the real people and the real place they happened".

Mr Chong-Wen Wu, Chief HR officer at ASUS TeK computer Inc.

HRD performance indicators

coaching

talent

innovation

learning

outcome

on the job learning

e-learning

347

tifa

Thank you!

Peter Vandenbruaene

Learning Solution Manager

Training Institute of the Federal Administration
Bischoffsheimlaan 15, 1000 Brussels, BELGIUM
Ph: +32 2 229 73 60
peter.vandenbruaene@ofoifa.fgov.be

Service public fédéral
Personnel et Organisation
Federale Overheidsdienst
Personeel en Organisatie

.be

Discussant's Manuscript on Prof John Wanna's session on Performance Reporting and Performance Management discussed by Peter Vandenbruaene,

Learning Solution Manager, Training Institute of the Belgian Federal Administration.

In what follows, I would like to share with you my personal views on some of Prof Wanna's statements. I have tried to look at them in the context of my professional backgroup, which is Human Resource Development within the Belgian Federal Civil service. I do not claim that I have a lot of answers. But sometimes, sharing the same questions is already a first step on the road towards an answer.

Some questions I have asked myself, are:

1. Which performance indicators do we use within the field of Human resource development?
2. Is it possible that these indicators, as Prof Wanna suggests, are sometimes "distorting" and even "perverse"?
3. If so, are there any alternatives?

1. Which performance indicators do we use within the field of Human resource development?

Our Training Institute of the Belgian Federal Administration delivers training courses to all Belgian federal civil servants, through an open offer of catalogue trainings and tailor-made learning projects. For decades we have been measuring training activities within the civil service using different kinds of performance indicators.

Each year, target performance indicators are fixed and consequently lots of data are collected. We measure indicators such as the total number of civil servants trained per year (specified per ministry or agenccy), the total number of training days, the number of courses on offer, or even the number of hours of instruction or training inidividual civil servants have to get over a certain period of time. Sheets and sheets of excelfiles are used to collect and analyse these data. The rationale behind this is probably the idea that the more courses we offer, the more trainings days we organize and the more hours civil servants spend in classrooms, the more they will learn, the better they will perform, and the better their agencies will perform when dealing with the public. But is that so? Is their really a link between the number of hours spent on training and, let's say, the number or satisfied customers a ministry has, or the number of files treated per year by its staff?

On another level, we are comparing the total federal training investment as a whole with that of other organizations in the public or even private sector.

And finally, all our courses are evaluated. Or so we think. We always hand out evaluation forms to participants, asking them whether or not they liked the course, which comments they have; we

even ask them wether or not they will apply what they learnt in their daily working context. If a majority of them answers positively, we are thrilled and we spend our nights dreaming that they will actually do so. But will they? And even if they do, will there be any noticeable impact on the strategic goals of their organizations? Will the public at large who is the client of the civil service eventually notice any difference?

2. Might these indicators, as Prof Wanna suggests, sometimes be "distorting" and even "perverse"?

In fact, we are less and less sure nowadays that the peformance indicators we are currently using within our HRD context, are actually an indication of what really matters.

Could they be "distorting" and even "perverse" as Prof Wanna suggests?

I am afraid they are. Take the "total number of courses organized" for instance. With virtually no effort at all, and no real extra costs, we could easily obtain a 50% increase in the number of training courses offered just by splitting up each 4 day course into two separate 2-day courses.

As far as Total training investment is concerned, we feel so proud that it thas doubled over th elast ten years , and that we are now at a figure of about 2.9% of the total mass of salary. The higher that percentage, the happier we seem , because we feel that this indicator shows us how well we are stimulating learning within the federal adminsitration. But does it? What if one the reasons that training investment has risen significantly is simply that external consultant prices have increased? We might even prefer to spend more on useless training, since the more money we spend, the better we will be in international benchmarks on training investment!

If the number of hours of training offered is the indicator, we could well increase this number by sending more and more people to training courses. But at what cost? And to what effect?

Even if we take into account the indicators which result from training evaluation, results may not always indicate what they seem to. Results on "Happy sheets" (training evaluation in which participants answer a questionnaire usually at the end of training) may provide us with some information on participants' having a really great time together and really enjoying the trainer (or the great hotel facilities where the training took place), they guarantee in no way that the training will have any lasting impact.

3. Are there any alternatives?

At the Belgian Training Institute of the federal Administration, we believe that two of the most

important elements that enable us to create more realistic, context-driven performance indicators in the field of HRD are conducting good Training Needs Analysis (TNA) and paying a lot of attention to good Training Evaluation (TE). These are not two isolated items, in fact, they go hand in hand. A good TNA is a basic requirement for good evaluation process.

What is "good Training Needs Analysis" and what is "good evaluation"?

We do not have the magic aswer. But what we have found so far is that a four level GAP-analysis helps us in the process of conduction a training needs analysis. What we are trying to do is not new, nor is it shocking: we are trying to compare the current situation ("as is") with the future, ideal situation ('to be') and this on four different levels, corresponding to the four levels of the Kirkpatrick's model for Training needs analysis: the organisation as a whole (where 'impact' should be measured), the day-to-day working environment (measuring 'transfer'), the level of individual competences (where we can evaluate wether there has been an increase in competences and/or knowledge) and the level of training course itself (the 'happy sheet'-level).

This four level approach has its impact on the way our HRD performance indicators should be defined. Instead of creating batteries of indicators measuring activities going on at the lowest level ("what takes place in class and how this is perceived by participants"), we had better define them at the third or fourth level of Kirkpatrick's evaluation model.

Last week, and thanks to the Taiwanese National Academy of the Civil Service, I was fortunate to attend a lecture on Talent Development by Mr Wu, Chief HR officer at ASUS TeK computer Inc. What he mentioned about Key Performance Indicators regarding Talent Development struck me as being so to the point: "it's not about the number of people taking a course, it is all about comparing interventions that should take place at the right time, for the right people at the right place with the real time, the real people and the real place they happened". To me, this is a fine example of how, by integrating Human resources into the business of the organisation, performance management can be lifted to a much more effective level.

Conclusion and challenges

Does this mean we should stop measuring the number of trainings organized, the number of hours of instruction or civil servants are getting or participants' reactions to a training course as noted down in happy sheets?

Probably not. But we should be very careful when drawing conclusions from the data thus obtained.

And we should be ready for the future. Instead of the traditional "teacher-student" approach,

where trainers help our trainees in a classroom context, we are shifting towards a learning context, where individuals or groups become responsible for their learning. The generation Y entering our civil service is alredy used to learning in a rapidly changing context, making use of a variety of alternative learning tools and learning styles. Are the performance indicators we are using capable of measuring that new way of learning as well? If not, we might find a significant decrease in terms of training hours and courses taught, whereas in reality a lot more learning is going on.

We are all attending this HR conference and we are spending a lot of time here. Time is precious. Time is money. And yet we are all here because we hope that attending this conference will enable us to learn something. So there is a lot of learning going on here. But tomorrow we will be back in our offices. Which performance indicator would we like to have to truly measure the impact of this event on us, our or working units, on our organizations? Are we only going to write down an extra number in the excel column that says "number of congresses and events taken in 2012"? Or do we need another framework altogether, and think rather of a performance indicator that takes into account where we are now and where we would like to be in the future and how this congres can be a bridge that helps us going there?

The fact that we are asking this question together, may be the first step towards an answer⋯