



***INDICATEURS FINANCIERS ET
INDICATEURS NON FINANCIERS.
IMPORTANCE DES INDICATEURS
FINANCIERS, ET ORDRE DE
PRESENTATION DES INDICATEURS :
EFFETS SUR L'EVALUATION DE LA
PERFORMANCE DES MANAGERS***

Eric CAUVIN^a, Bruce R. NEUMANN^b et Michael L. ROBERTS^c

^a Maître de Conférences

CEROG-IAE Aix en Provence

Professeur Affilié à l'International University of Monaco

^b Professeur, Business School

University of Colorado at Denver and Health Sciences Center

^c Professeur Associé, Business School

University of Colorado at Denver and Health Sciences Center

Le débat sur l'importance des indicateurs non financiers et financiers dans le processus d'évaluation des performances est loin d'être clos. L'objectif de notre recherche est d'examiner, de façon expérimentale, les facteurs qui influencent le choix des évaluateurs en faveur des indicateurs financiers ou non financiers dans l'évaluation des performances d'une entreprise. Nous avons ainsi d'une part manipulé l'ordre de présentation des indicateurs financiers et non financiers dans le but de tester si les limitations cognitives conduisaient les évaluateurs à considérer seulement les premiers indicateurs qui leur étaient présentés. D'autre part, nous avons manipulé l'importance relative perçue des indicateurs financiers afin d'examiner si le biais en faveur des indicateurs financiers, souvent démontré, pouvait être expliqué par la nature même de ces indicateurs. L'importance relative perçue des indicateurs financiers se fondait sur les résultats d'une autre étude que nous avons conduite sur un échantillon indépendant.

Debate on the weight of financial vs. non financial indicators in performance measurement process is still going on. The purpose of this research is to examine, experimentally, factors that influence whether corporate managers exhibit bias toward financial versus nonfinancial corporate performance measures. First, we manipulated presentation order of financial and nonfinancial measures to test whether cognitive processing limits result in decision makers placing more weight on the first few items presented. Second, we manipulated the strength of the financial performance measures (the perceived relative importance based on rankings by an independent sample) to examine whether prior mixed results of financial measures bias among corporate managers can be explained by differences in the relative strength or importance of the particular set of financial and nonfinancial measures.

Mots clés : Evaluation des performances, Indicateurs financiers vs. Non financiers, Biais cognitif, Ordre de présentation des informations ; Performance measurement, Financial indicators vs. NonFinancial indicators, Cognitive bias, Primacy effect.

L'objectif de notre recherche était d'examiner, de façon expérimentale, les facteurs qui influencent le choix des évaluateurs en faveur des indicateurs financiers ou non financiers dans l'évaluation des performances d'une entreprise.

En dépit de l'importance accordée aux indicateurs non financiers dans les modèles d'évaluation des performances (Kaplan et Norton 1996), Ittner et al. (2003) ont montré que dans la réalité les managers tendaient à ignorer ces mesures souvent considérées comme trop subjectives et trop manipulables. Plus récemment, les résultats d'une étude expérimentale ont montré que dans leurs *attitudes* les évaluateurs considéraient les indicateurs financiers comme plus pertinents, plus fiables et plus comparables que les indicateurs non financiers. En revanche, dans leurs *jugements* ce biais disparaissait, et il apparaissait même que les évaluateurs les plus expérimentés accordaient plus d'importance aux indicateurs non financiers (Cauvin et al. 2007).

Se fondant sur les travaux des psychologues Slovic et MacPhillamy (1974), Lipe and Salterio (2000), ont montré que dans l'évaluation comparative de la performance des managers, la sur

charge cognitive conduisait les évaluateurs à privilégier un nombre plus réduit d'indicateurs ; en l'occurrence les indicateurs que les managers avaient en commun. Une étude plus récente dans le même domaine de la psychologie a révélé qu'un problème de surcharge cognitive pourrait avoir des implications sur l'analyse des informations. En effet, Halford et al. (2005) ont montré qu'un être humain ne pouvait mentalement analyser plus de quatre variables simultanément au risque d'affaiblir la qualité de son jugement. Or les systèmes d'évaluation des performances telles que le Balanced Scorecard proposent généralement entre seize et vingt quatre indicateurs. Aussi, le processus d'évaluation des performances qui s'inscrit dans une démarche holistique, dans laquelle les décideurs analysent simultanément toutes les informations, pourrait faire l'objet de limitations inconscientes liées au nombre réduit d'informations analysées.

Dans notre recherche, nous avons modifié le cas expérimental utilisé par Lipe et Salterio (2000), Banker et al. (2004), ainsi que Libby et al. (2004), où il s'agissait d'évaluer la performance des deux managers de deux divisions d'une entreprise.

Nous avons d'une part manipulé l'ordre de présentation des indicateurs financiers et non financiers dans le but de tester, comme l'ont montré Halford et al. (2005), si les limitations cognitives conduisaient les évaluateurs à considérer seulement les premiers indicateurs qui leur étaient présentés. D'autre part, nous avons manipulé l'importance relative perçue des indicateurs financiers afin d'examiner si le biais en faveur des indicateurs financiers démontré précédemment (Schiff et Hoffman 1996 ; Ittner et al. 2004 ; Cauvin et al. 2007) pouvait être expliqué par la nature même de ces indicateurs. L'importance relative perçue des indicateurs financiers s'appuyait sur les résultats d'une autre étude conduite sur un échantillon indépendant.

La revue de la littérature fondant notre étude expérimentale est présentée dans la partie suivante. Les autres parties s'intéressent respectivement à la méthodologie suivie, aux résultats obtenus et aux conclusions.

1. Revue de la littérature

Plusieurs travaux de recherche récents se sont intéressés à l'évaluation de la performance des managers. Malina et Selto (2001) ont montré qu'une évaluation qui combine des informations financières et non financières se révèle efficace dans le contrôle de la mise en œuvre d'une stratégie. Néanmoins les auteurs ont conclu que pour être un processus de contrôle efficace, les évaluations devaient être perçues comme précises, objectives et vérifiables. Des tensions et des conflits entre évaluateurs et évalués ont même pu être observés dès lors que les indicateurs de performance étaient perçus comme subjectifs (Malina et Selto 2001). La subjectivité perçue des indicateurs peut même conduire au rejet des systèmes d'évaluation des performances composés de mesures financières et non financières, et au retour de démarches plus centrées sur les données financières seulement (Ittner et al. 2003).

Se fondant sur les travaux de Lipe et Salterio (2000), plusieurs études expérimentales récentes ont été conduites sur l'utilisation du Balanced Scorecard (Kaplan et Norton 1996) dans l'évaluation de la performance des managers de différentes divisions d'une même entreprise. Lipe et Salterio ont en effet montré que les évaluateurs privilégient les indicateurs de performance communs aux deux divisions et ignorent les indicateurs qui leur sont uniques. Ils ont ainsi montré que la moitié des informations contenues dans le Balanced Scorecard étaient ignorées. En effet, le Balanced Scorecard implique généralement l'appréciation d'une vingtaine d'indicateurs¹ différents avant de porter un jugement holistique. Lipe et Salterio expliquent le biais en faveur des mesures communes en s'appuyant sur les travaux menés par les psychologues Slovic et MacPhillamy (1974) qui ont montré que les évaluateurs cherchaient à réduire la charge cognitive d'analyse des informations en sélectionnant celles qui étaient communes aux évalués.

Les résultats d'autres études ont montré que le biais en faveur des mesures communes pouvait être réduit, mais pas totalement éliminé, en communiquant sur l'importance stratégique des indicateurs uniques (Banker et al. 2004), en augmentant leur objectivité ou en attestant de leur fiabilité (Libby et al. 2004), et par l'entraînement (Dilla and Steinbart 2005). Roberts et al. (2004) soutiennent que la seule méthode susceptible d'éliminer complètement le biais en faveur des

¹ Les études citées proposent des Balanced Scorecard construits autour de seize à vingt huit indicateurs.

mesures communes est la désagrégation de l'évaluation holistique de plusieurs indicateurs en une série d'évaluations portant sur chaque indicateur.

Les travaux en matière de prise de décisions ont montré l'utilisation de différentes stratégies (Abelson et Levi 1985), dont la stratégie coût-bénéfice, dans le but de réduire l'effort cognitif. Le coût est l'effort nécessaire pour prendre une décision, le bénéfice représente la meilleure alternative potentielle (Beach et Mitchell 1978 ; Russo et Doshier 1983). Le choix par attributs constitue l'une des stratégies coût-bénéfice. Le décideur sélectionne un critère (habituellement le plus important), classe chaque alternative d'après ce critère, puis passe au critère suivant le plus important (Tversky 1969 ; Bettman et Kakkar 1977). Les évaluateurs peuvent être conduits à réduire leur effort cognitif en sélectionnant un échantillon d'indicateurs (Payne et al. 1988 ; Todd et Benbasat 1994 ; Lipe et Salterio 2000). Ce postulat s'explique aussi par les résultats de récents travaux qui montrent que l'être humain est capable d'analyser au maximum quatre variables à la fois. Au-delà de quatre variables la qualité de l'analyse relève du hasard (Halford et al. 2005). Néanmoins, on ne peut pas forcément attribuer au biais des indicateurs financiers, noté par Ittner et al. (2003) et suggéré par Malina et Selto (2001), le besoin de réduire l'effort cognitif en réduisant le nombre d'informations à analyser. Dans leurs études, c'est la recherche d'une plus grande objectivité qui est mentionnée.

Trois études expérimentales récentes ont examiné l'utilisation relative d'indicateurs financiers et non financiers dans la construction d'un système d'évaluation des performances. Schiff et Hoffman (1996) et Krumwiede et al. (2005) ont montré que les indicateurs financiers étaient privilégiés lorsque l'évaluation de la performance portait sur un domaine d'activité, mais que c'étaient les indicateurs non financiers qui étaient préférés dès lors que la performance des managers devait être appréciée. Luft et Shields (2001) ont par ailleurs montré que l'utilisation d'indicateurs non financiers était privilégiée lorsqu'il s'agissait de prévoir les performances financières futures. Ces études expérimentales se sont donc principalement intéressées à l'importance relative des indicateurs financiers et non financiers.

Cauvin et al. (2007), ont montré qu'un biais en faveur des indicateurs financiers² existait dans l'attitude des managers vis-à-vis des critères comptables de pertinence, de fiabilité et de comparabilité (FASB 1980, 2005) des différents types d'indicateurs, mais que dans leur jugement cette préférence n'était pas exercée. Au contraire, leur étude montre que les managers les plus expérimentés accordent plus d'importance aux mesures non financières. Cauvin et al. (2007) présente un modèle des pistes causales qui montre que l'évaluation de deux divisions d'une même entreprise est fonction du : (1) du type d'indicateur (financier ou non financier) qui est préféré qui est en fait expliqué par (2) les critères comptables (pertinence, fiabilité et comparabilité) attachés aux indicateurs et par (3) l'expérience des managers en matière d'évaluation. Leurs résultats peuvent expliquer les constats d'études antérieures (e.g., Ittner et al. 2003 ; Malina et Selto 2001) dans lesquelles les évaluateurs les plus expérimentés semblaient utiliser des indicateurs non financiers pour évaluer leurs collaborateurs même si ceux-ci sont généralement perçus comme moins pertinents, moins fiables, moins comparables et moins importants.

Se fondant sur ce corpus théorique, l'objectif de notre étude expérimentale était de tester les hypothèses suivantes.

D'abord, dans le but de réduire l'effort cognitif les évaluateurs ont tendance à sélectionner un nombre limité d'informations à analyser (Payne et al. 1988 ; Todd et Benbasat 1994 ; Lipe et Salterio 2000). Cette stratégie peut par conséquent engendrer la prise en compte inconsciente de l'ordre de présentation des informations (order effect). En effet, lorsque les évaluateurs sont face à un tableau de bord contenant plus de quatre indicateurs financiers et non financiers un effet de primauté (primacy effect) peut influencer le poids relatif placé sur les indicateurs listés en premier quelle que soit leur nature (financier ou non financier).

D'où la formulation d'une première hypothèse :

² Merriam-Webster définit un biais comme une tendance, un préjugé (Merriam-Webster 1998). Le Petit Larousse définit un biais comme une direction oblique par rapport à une direction principale (Petit Larousse 1998). Nous utilisons ce terme pour indiquer une préférence en faveur des indicateurs financiers en dépit du fait que l'équilibre entre les indicateurs financiers et non financiers devrait s'imposer en fonction de la stratégie (Kaplan et Norton 1996 ; Banker et al. 2004).

H1 : un effet de primauté (primacy effect) sur les premiers indicateurs présentés, qu'ils soient financiers ou non financiers, a un impact sur l'évaluation des performances des managers.

Ensuite, toujours dans le but de réduire le nombre d'informations à analyser, les évaluateurs vont plutôt privilégier les informations financières ou non financières selon l'importance relative qu'ils accordent à ces informations.

D'où la formulation d'une seconde hypothèse :

H2 : lorsque des informations financières auxquelles les évaluateurs accordent une importance relative forte sont proposées, les évaluateurs privilégient ces informations aux dépens d'informations non financières ; lorsque des informations non financières auxquelles les évaluateurs accordent une importance relative forte sont proposées, les évaluateurs privilégient ces informations aux dépens d'informations financières.

Enfin, nous avons considéré que le critère d'ordre d'apparition des informations (H1) combiné au critère d'importance relative accordée par les évaluateurs à des informations, qu'elles soient financières ou non financières (H2), pouvaient cumuler leur effet respectif.

D'où la formulation d'une troisième hypothèse :

H3a : les évaluateurs surévaluent le manager dont les performances financières sont plus élevées par rapport au manager dont les performances non financières sont plus élevées lorsque :

- *les informations financières sont listées en premier et,*
- *les informations financières auxquelles les évaluateurs accordent une importance relative forte leur sont proposées.*

Plutôt que si une seule des deux conditions est remplie.

H3b : les évaluateurs surévaluent le manager dont les performances non financières sont plus élevées par rapport au manager dont les performances financières sont plus élevées lorsque :

- *les informations non financières sont listées en premier et,*

- *les informations financières auxquelles les évaluateurs accordent une importance relative faible leur sont proposées.*

Plutôt que si une seule des deux conditions est remplie.

2. Méthodologie

Les participants à l'étude devaient résoudre un cas fondé sur l'évaluation de deux unités géographiques de l'entreprise RadWear, une division de WCS Incorporated, une entreprise de détail spécialisée dans les vêtements féminins. Ce cas est adapté de celui utilisé par Lipe et Salterio (2000), qui avait été lui-même adapté du cas de mise en œuvre du Balanced Scorecard utilisé par Kaplan et Norton (1996) : Kenyon Stores. Un extrait en anglais est proposé en Annexe 1. Les participants devaient assumer le rôle du manager de division RadWear dont la mission est la vente de vêtements originaux sur le marché du prêt-à-porter réservé aux adolescents. Les participants avaient à leur disposition un tableau annuel d'évaluation des performances des deux managers régionaux ainsi que les objectifs qui leur avaient été fixés en début d'année. Le tableau contenait huit indicateurs proposés sans ordre précis ; la moitié était des indicateurs financiers et l'autre moitié des indicateurs non financiers (Cf. Annexe 2).

Kaplan et Norton (1996) suggèrent l'utilisation de quatre catégories d'indicateurs : les indicateurs financiers, les indicateurs liés aux clients, les indicateurs liés aux processus internes et les indicateurs liés à l'apprentissage organisationnel. Néanmoins, ces quatre axes peuvent aussi être classés en deux catégories : les indicateurs financiers et les indicateurs non financiers. Ainsi, dans les tableaux de bord utilisés par Lipe et Salterio (2000), les indicateurs financiers ainsi que les indicateurs liés aux processus internes représentent les indicateurs traditionnellement appelés financiers. En revanche, les indicateurs liés aux clients ainsi que ceux liés à l'apprentissage organisationnel représentent les indicateurs non financiers. Dans les recherches précédentes conduites sur le Balanced Scorecard (e.g., Lipe et Salterio 2000, 2002 ; Banker et al. 2004 ; Libby et al. 2004 ; Roberts et al. 2004), la classification des indicateurs financiers et non financiers reste simple. Nous avons choisi huit mesures parmi celles utilisées par Lipe et Salterio

(2000) et nous avons classé ces huit mesures en tant que mesures financières ou non financières (Cf. Annexe 1).

Afin de tester la validité de nos classifications, nous avons proposé une liste d'indicateurs non classés à vingt neuf étudiants de l'Université du Colorado à Denver (USA) de niveau graduate en gestion³, et leur avons demandé de classer chaque indicateur comme financier (codé 1) ou non financier (codé 0). La moyenne de la somme de ces classements est de 3,45 (écart type : 0,57) pour les indicateurs financiers, et 1,14 (écart type : 0,95) pour les indicateurs non financiers. La différence (2,31 ; écart type : 1,07) est significative ($p < 0,0001$; $t = 11,60$). Les classifications indépendantes confirment donc nos classifications a priori.

Chaque tableau de bord contenait huit indicateurs séparés. Les indicateurs étaient les mêmes pour chacune des deux régions ce qui démarque notre étude des récents travaux sur les effets des indicateurs communs et uniques (e.g., Lipe et Salterio 2000 ; Lipe et Salterio 2002 ; Banker et al. 2004 ; Libby et al. 2004 ; Roberts et al. 2004). Comme dans l'étude de Lipe et Salterio (2000), les deux régions avait dépassé leurs objectifs sur tous les indicateurs.

Les participants devaient évaluer la performance de chaque manager régional sur une échelle de 0 (Très faible) à 100 (Très élevée). Cette échelle est celle utilisée par Lipe et Salterio (2000). A la suite de cette évaluation, les participants devaient répondre à des questions sur l'importance de chacun des indicateurs ; sur la pertinence, la fiabilité, et la comparabilité des indicateurs financiers comparés aux indicateurs non financiers ; sur la difficulté, le réalisme et la compréhensibilité du cas ; et à des questions démographiques.

Quatre vingt douze étudiants en gestion de l'Université du Colorado à Denver (USA) ont participé à notre étude. 65% étaient des étudiants de MBA, 2% étaient étudiants en MS Comptabilité, 30% appartenaient à d'autres MS en business (principalement en Finance) et 13% étaient inscrits en MBA et MS. L'âge moyen des participants était 29 ans, et l'expérience professionnelle moyenne était 6 ans. 55% des participants étaient des hommes. 60% ont déclaré avoir déjà participé à un processus d'évaluation de leurs subordonnés dans le cadre de leur

³ A noter que ces étudiants ne participent pas à notre étude principale.

activité professionnelle. Comme dans l'étude menée par Lipe et Salterio (2000), le cas a été résolu en classe.

Les indicateurs financiers en tête de liste ont été présentés à quarante cinq participants ; les indicateurs non financiers ont été présentés en tête de liste aux quarante sept autres participants. Cinquante quatre participants ont reçu des indicateurs financiers plus importants en plus des indicateurs non financiers ; trente huit participants ont reçu des indicateurs financiers moins importants.

Les variables dépendantes sont l'évaluation de la performance (de 0 à 100) du manager de la région 1, qui excellait sur les indicateurs financiers, et l'évaluation du manager de la région 2 qui excellait sur les indicateurs non financiers. Nous avons également utilisé la différence d'évaluation, c'est-à-dire le score du manager de la région 1 moins le score du manager de la région 2.

Globalement, les participants considèrent que le cas est facile à comprendre, n'est pas difficile à résoudre, et est réaliste. Sur la base d'une échelle de -5, pas du tout d'accord, à 5, tout à fait d'accord, le score moyen pour le réalisme est de 1,4 ; le score moyen pour la compréhensibilité est de 2,7, et le score moyen pour la difficulté est de -2,5. Toutes ces moyennes sont significativement différentes de zéro ($p < 0,0001$). Ces questions de contrôle sont les mêmes que celles utilisées par Lipe et Salterio (2000). Les participants ont confirmé que le rapport d'évaluation des performances de RadWear contenait à la fois des informations financières et non financières (3,9, $p < 0,0001$), que le manager de la région 1 avait réalisé une meilleure performance sur les indicateurs financiers que sur les indicateurs non financiers (3,4, $p < 0,0001$), et ont infirmé que le manager de la région 2 avait réalisé une meilleure performance sur les indicateurs financiers que sur les indicateurs non financiers (-2,7, $p < 0,0001$). Les participants devaient indiquer parmi les huit indicateurs de performance utilisés par RadWear lesquels étaient considérés comme importants dans leur évaluation. Les participants ayant reçu les indicateurs financiers les plus importants (relativement) ont évalué trois indicateurs financiers comme très importants dans leur évaluation. Les participants n'ayant pas reçu d'indicateurs financiers importants ont choisi un seul indicateur financier comme très important dans leur évaluation.

3. Les résultats

Les résultats des deux managers régionaux selon une analyse de la variance (ANOVA) sont présentés dans le Tableau 1.

Tableau 1
Indicateurs financiers vs. Indicateurs non financiers, Importance relative des indicateurs financiers, et Effet de l'ordre sur l'évaluation de la performance de chaque manager

<u>Source</u>	<u>DF</u>	<u>Type III SS</u>	<u>Mean Square</u>	<u>F Value</u>	<u>Pr > F</u>
Ordre	1	52,80	52,80	0,34	0,56
ImportIndFin	1	3,71	3,71	0,02	0,88
Ordre* ImportIndFin	1	179,99	179,99	1,17	0,28
Erreur	88	13 491,58	153,31		

<u>Source</u>	<u>DF</u>	<u>Type III SS</u>	<u>Mean Square</u>	<u>F Value</u>	<u>Pr > F</u>
FiNon	1	14,02	14,02	0,24	0,63
FiNon*Ordre	1	290,99	290,99	4,98	0,01
FiNon* ImportIndFin	1	216,24	216,24	3,70	0,03
FiNon*Ordre * ImportIndFin	1	24,07	24,07	0,41	0,52
Erreur	88	5 143,13	58,45		

ImportIndFin renvoie au test des indicateurs financiers relativement importants tels que : Net Income Growth et ROI versus les indicateurs financiers les moins importants tels que : Return on Sales et Market Share (in sales dollars). Les indicateurs non financiers étaient les mêmes dans les deux conditions.

Ordre indique que les quatre indicateurs financiers étaient présentés avant ou après les quatre indicateurs non financiers.

FiNon renvoie au test où le manager de la région 1 excelle sur les performances financières et le manager de la région 2 excelle sur les performances non financières.

Les variables dépendantes de l'ANOVA sont les évaluations des deux managers régionaux, région 1 et région 2, où le manager de la région 1 excelle sur les performances financières et le manager de la région 2 excelle sur les performances non financières.

Dans un premier temps, nous avons testé si l'ordre de présentation des informations ou la présentation d'informations financières perçues comme relativement importantes affectaient l'évaluation des deux managers. Les deux tests sont significatifs ($p < 0,05$). D'une part, le test « performance financière versus non financière et l'ordre » est significatif ($p = 0,01$), ce qui indique que les deux managers régionaux ont été évalués différemment selon l'ordre de présentation des informations financières et non financières. D'autre part, le test « performance financière versus non financière et l'importance relative des indicateurs financiers » est significatif ($p = 0,03$). Ceci indique que les deux managers régionaux ont été évalués différemment selon que les participants possédaient les indicateurs financiers relativement importants ou relativement faibles.

Dans le but de confirmer ces constats, nous avons testé l'impact de l'ordre de présentation des informations et de l'importance relative des indicateurs financiers sur la différence calculée entre le score du manager 1 et le score du manager 2.

Tableau 2

Analyse de la variance et Tests-t sur l'impact de l'effet de l'ordre de présentation et l'importance relative des indicateurs financiers sur les différences dans l'évaluation de la performance des managers

Panel a :

<u>Source</u>	<u>DF</u>	<u>Type III SS</u>	<u>Mean Square</u>	<u>F Value</u>	<u>Pr > F</u>
Ordre	1	432,48	432,48	4,98	0,01
ImportIndFin	1	581,98	581,98	3,70	0,03
Ordre* ImportIndFin	1	48,13	48,13	0,41	0,52

Erreur	88	10 286,25	116,89
Total	91	11 295,65	

Panel b : t-tests

Ordre :	Moyennes (écarts types.)		
	<u>Manager Region 1</u>	<u>Manager Region 2</u>	<u>Diff.</u>
Indicateurs financiers présentés en premier	+2,2 (11,4)	-2,6 (10,5)	4,6*
Indicateurs non financiers présentés en premier			
Importance des indicateurs financiers :			
Indicateurs financiers importants	+1,6 (10,2)	-2,8 (12,1)	4,3*
Indicateurs financiers peu importants			

* $p < 0,05$

Ordre indique si les quatre indicateurs financiers étaient présentés avant ou après les quatre indicateurs non financiers.

ImportIndFin renvoie au test des indicateurs financiers relativement importants tels que : Net Income Growth et ROI versus les indicateurs financiers les moins importants tels que : Return on Sales et Market Share (in sales dollars). Les indicateurs non financiers étaient les mêmes dans les deux conditions.

La variable dépendante de l'ANOVA est la différence dans les évaluations des deux managers régionaux, région 1 moins région 2, où le manager de la région 1 excelle sur les performances financières et le manager de la région 2 excelle sur les performances non financières.

Le Tableau 2 (panel a) montre que l'ordre de présentation des informations ($f = 4,98$, $p = 0,01$) et de l'importance relative des indicateurs financiers ($f = 3,70$, $p = 0,03$) expliquent de façon significative les différences dans l'évaluation des managers des deux régions.

Le panel b du Tableau 2 montre que les tests-t paramétriques conduits sur les différences moyennes des évaluations des managers sont également significatifs. Lorsque les mesures financières sont présentées en premier, le manager de la région 1 est préféré (+2,2) ; lorsque les mesures non financières sont présentées en premier, le manager de la région 2 est préféré (-2,6).

S'agissant de l'ordre, la différence entre les deux groupes (4,6) est significative ($t = 2,09$, $p = 0,02$). Lorsque les mesures financières sont relativement plus importantes, le manager de la région 1 est préféré (+1,6) ; lorsque les mesures financières sont relativement moins importantes, le manager de la région 2 est préféré (-2,8). S'agissant de l'importance relative des mesures financières, la différence entre les deux groupes (4,3) est significative ($t = 1,9$, $p = 0,03$).

Les différences moyennes des évaluations du manager 1, plus performant sur les indicateurs financiers, et des évaluations du manager 2, plus performant sur les indicateurs non financiers, pour chacune des quatre expérimentations apparaissent au Tableau 3.

Tableau 3

Les différences moyennes dans l'évaluation de la performance des managers (i.e., Evaluation de la Région 1 moins Evaluation de la Région 2) selon les groupes définis par l'importance relative des indicateurs financiers et l'ordre de présentation des indicateurs

Groupes	N	Différence			
		<u>moyenne</u>	<u>Ecart type.</u>	<u>Minimum</u>	<u>Maximum</u>
IndFinImpH, FinPrem	26	+3,46*	10,47	-20,00	+30,00
IndFinImpH, NonFinPrem	28	0,18	9,76	-20,00	+15,00
IndFinImpF, FinPrem	19	0,53	12,68	-35,00	+25,00
IndFinImpF, NonFinPrem	19	-6,05*	10,75	-30,00	+15,00

* $p < 0,05$

ImportIndFin renvoie au test des indicateurs financiers relativement importants tels que : Net Income Growth et ROI versus les indicateurs financiers les moins importants tels que : Return

on Sales et Market Share (in sales dollars). Les indicateurs non financiers étaient les mêmes dans les deux conditions.

FinPrem/NonFinPrem indique que les quatre indicateurs financiers étaient présentés avant ou après les quatre indicateurs non financiers.

La différence moyenne, minimum et maximum renvoie au test dans lequel le manager de la région 1 excellait sur les indicateurs financiers et le manager de la région 2 excellait sur les indicateurs non financiers. Les différences moyennes montrent le score du manager de la région 1 moins le score du manager de la région 2

Lorsque les mesures financières sont présentées en premier et que les mesures financières ont une importance relative élevée, le manager de la région 1 est surévalué par rapport au manager de la région 2 (+3,46, $p < 0,05$). Lorsque les mesures non financières sont présentées en premier et que les mesures financières ont une importance relative faible, le manager de la région 2 est surévalué par rapport au manager de la région 1 (-6,05, $p < 0,05$). En revanche, lorsque les mesures non financières sont présentées en premier et que les mesures financières ont une importance relative élevée, il n'y a pas de différence dans l'évaluation des deux managers (0,18, $p > 0,10$). De même, lorsque les mesures financières sont présentées en premier et que les mesures financières ont une importance relative faible, il n'y a pas de différence dans l'évaluation des deux managers (0,53, $p > 0,10$).

CONCLUSIONS

Cette étude s'inscrit dans la continuité des recherches antérieures qui ont montré d'une part que selon qu'il s'agissait d'évaluer une business unit ou un manager les indicateurs financiers ou non financiers avaient la préférence des évaluateurs (Schiff et Hoffman 1996), et d'autre part que les évaluateurs considèrent que le poids des indicateurs financiers était plus élevé que celui des indicateurs non financiers lorsqu'il s'agissait d'évaluer la performance d'un subordonné (Ittner et al. 2003). Nous nous intéressons également au paradoxe mis en évidence dans nos précédents travaux : les indicateurs financiers sont considérés comme plus pertinents, plus fiables, plus comparables, et globalement, plus importants que les indicateurs non financiers, en dépit du fait

que les indicateurs financiers ne sont pas surévalués par rapport aux indicateurs non financiers dans les évaluations (Cauvin et al. 2007).

Nos résultats montrent que les évaluations de la performance fondées sur des indicateurs financiers et non financiers sont affectées à la fois par l'ordre de leur présentation et par l'importance relative (ou non importance) des indicateurs financiers proposés. Selon que les indicateurs financiers ou que les indicateurs non financiers sont présentés en premier, la performance du manager qui excelle sur ces types de mesure est évaluée de façon significativement plus élevée. Notre hypothèse H1 est donc confirmée (un effet de primauté (primacy effect) sur les premiers indicateurs présentés, qu'ils soient financiers ou non financiers, a un impact sur l'évaluation des performances des managers. Lorsque les indicateurs de performance incluent des indicateurs financiers relativement importants plutôt que relativement peu importants, la performance du manager qui excelle sur les mesures financières est surévaluée. Notre hypothèse H2 est donc confirmée (lorsque des informations financières auxquelles les évaluateurs accordent une importance relative forte sont proposées, les évaluateurs privilégient ces informations aux dépens d'informations non financières ; lorsque des informations non financières auxquelles les évaluateurs accordent une importance relative forte sont proposées, les évaluateurs privilégient ces informations aux dépens d'informations financières).

Nos résultats confirment également nos hypothèses H3a et H3b fondées sur la combinaison de l'effet de primauté et de l'importance relative des indicateurs. Les évaluateurs accordent un score plus élevé au manager de la région 1, qui excelle dans les mesures financières, lorsque (a) les indicateurs financiers sont présentés en premier et (b) lorsque les indicateurs financiers sont relativement plus importants (H3a) plutôt que lorsqu'une seule des deux conditions est remplie (H3a validée), et les évaluateurs accordent un score plus important au manager de la région 2, qui excelle dans les indicateurs non financiers, lorsque (a) les indicateurs non financiers sont présentés en premier et (b) lorsque les indicateurs financiers ont une importance relative faible (H3b) plutôt que lorsqu'une seule des deux conditions est remplie (H3b validée).

Les résultats suggèrent également que les limites cognitives de la capacité d'analyse de l'être humain (Halford et al. 2005) peuvent avoir des conséquences involontaires sur l'évaluation des

performances dans laquelle plus de quatre indicateurs doivent être analysés pour parvenir à une évaluation globale. Nous pensions, et notre étude le confirme, que la première catégorie d'indicateurs présentée (soit quatre indicateurs) aurait un plus grand impact sur le processus d'évaluation que ne l'auraient les mesures présentées par la suite.

La décision des évaluateurs d'accorder plus ou moins d'importance aux différents indicateurs de performance qui leur sont proposés montre la complexité du processus. Nos résultats montrent trois raisons pour lesquelles les limites cognitives peuvent affecter leurs jugements. Les managers impliqués dans le processus d'évaluation des performances doivent être sensibles à trois facteurs, l'ordre de présentation des indicateurs, leur importance relative et le type de mesures, qui pourraient affecter leur jugement.

ANNEXES

Annexe 1 : extrait du cas proposé aux participants à l'étude

WCS Inc. Case

Background

Women's Clothing Store, Inc. is a firm specializing in the retailing of women's apparel. WCS has five subsidiaries, each directed at providing clothing to particular niches within the women's apparel market. WCS operates throughout North America and is currently considering a foray into the European markets.

WCS's largest subsidiary is RadWear. RadWear operates retail outlets specializing in apparel for the teenage girl. WCS operates its subsidiaries in a fairly decentralized manner. Each has a divisional manager who is responsible for the division's performance and for providing communication between WCS's central administration and his/her division. Each division, in turn, is organized into geographic operating regions.

Pat Jenks is the chief financial officer (CFO) of WCS Inc. In 2003, Pat attended a symposium at the Stanford Business School. Stanford professors introduced the symposium participants to a new management tool called Strategic Performance Evaluation containing a set of performance measures carefully chosen to represent important aspects of a business unit's financial performance, customer relations, internal business processes, and personnel development. These measures should be drivers of the unit's success and linked to its strategy and mission.

2003 was a rather lackluster year for WCS and Pat decided in early 2004 to try out the new strategic evaluation concept. Several steps were involved in doing this. First, Pat met with WCS's top management team to define the firm's overall mission. This team determined that the following mission statement was appropriately inspirational and captured the company's goal:

We will be an outstanding apparel supplier in each of the specialty niches served by WCS.

The top management team then met with each divisional manager to communicate this firm-wide mission and to discuss the division manager's role in developing a Strategic Performance Evaluation for his or her division. After the divisional evaluations were developed, each divisional manager met again with top management to explain his or her division's performance evaluation, to answer questions, and to make necessary adjustments as requested by top management. The performance evaluations were developed in time for pilot testing in the last quarter of fiscal year 2004. Based on WCS's experience in that quarter, the performance evaluations were adjusted for use in fiscal year 2005. Below is a description of RadWear's experience in developing and using the Strategic Performance Evaluation.

RadWear

When the concept of the Strategic Performance Evaluation was explained to Chris Peters, manager of the RadWear division, Chris was excited by the chance to specify the drivers of performance in the division. Chris believes that the Strategic Performance Evaluation idea has the potential to propel WCS and RadWear to vastly improved performance.

In order to develop a Strategic Performance Evaluation for RadWear, Chris first met with Jim Taylor and Bob Graham, the managers for the division's two largest geographical regions which produce 65 percent of the division's annual sales, and with the divisional management team. This team includes the divisional controller, marketing manager, head buyer, personnel manager, and manager of store operations. The management team described the target customer for RadWear and determined the factors they believe cause this customer to shop at RadWear, to return to RadWear, and to increase the percentage of her clothing purchased at RadWear. All of the resulting measures are controllable by RadWear's regional managers.

RadWear's staff was enthusiastic about the resulting Strategic Performance Evaluation and eager to use this new management tool in fiscal year 2005. RadWear's targeted Strategic Performance Evaluation is presented in Exhibit 1.

Exhibit 1

RadWear Divisional Strategic Performance Evaluation Measures and Targets for 2004

[Measures and targets omitted; see Table 1]

Performance in Fiscal Year 2005

The overall performance of WCS in fiscal year 2005 was good. Performance across divisions and regions, however, was somewhat uneven. The Strategic Performance Evaluation for RadWear's two largest regions, including actual amounts and percentages above targets for 2005, are provided in Exhibits 2 and 3, respectively.

Chris Peters now has the difficult task of evaluating the performance of the regional managers for 2005. As managers of the two largest regions, Jim Taylor and Bob Graham will receive the most attention. WCS and RadWear use the results of year-end reviews in several ways. They are used in determining merit raises and year-end bonuses. They are also used in decisions regarding promotion or movement within or out of the firm. Additionally, the performance reviews are used as a method for providing feedback and guidance to the regional managers regarding their performance and future actions.

Although Chris will do an initial evaluation, the final written review will take into account the discussion Chris will have with the individual divisional managers. Thus, the initial evaluation may be subsequently adjusted due to additional information provided by the evaluatees. Chris, however, believes it is important to do an initial evaluation before meeting with the managers in order to maintain objectivity and high standards; Chris believes later adjustments provide for equity based on the specifics of the regions' situations.

Your task

Please take the place of Chris Peters in evaluating the performance of Jim Taylor and Bob Graham, managers of Region 1 and Region 2, respectively. After reviewing their performance results in Exhibits 2 and 3, you should indicate an overall performance evaluation for each of the regional managers. The evaluation forms used by RadWear are attached (see pages 7 and 8).

Annexe 2 : rapport de performance de la Région 1, Indicateurs financiers relativement importants et relativement peu importants

Region 1 Strategic Performance Evaluation, Targets and Actuals for 2005

Measure	Actual	% better than target
Financial Measures (Stronger Measures)		
• Return on investment (2005 NI/2005 invested capital)	23.4%	17.14%
• Net income growth ([2005 NI - 2004 NI]/ 2004 NI)	18.3%	14.58%
• Sales growth ([2005 sales - 2004 sales]/2004 sales)	43.2%	23.33%
• Return on sales (2005 NI/2005 sales)	30%	25.00%
Financial Measures (Weaker Measures)		
• Return on sales (divisional NI/divisional sales)	28.1%	17.08%
• Market share relative to retail space (sales/total teen-wear retail square footage in 5 mile radius of stores)	\$91.7	14.63%
• Average markdowns (discounts)	12.2%	23.75%
• % Sales from new items designated as market leaders	31.1%	24.44%
Customer and Internal Business Measures (Both Conditions)		
• Customer satisfaction rating	88%	3.53%
• Mystery shopper program rating (Note 1 below)	86	1.18%
• % Returns to suppliers due to quality problems	5.8%	3.33%
• # Hours of employee training/employee	15.9	5.78%

Note 1: the Mystery Shopper program sends unidentified inspectors into stores. The inspectors are disguised as buyers and provide several ‘tests’ of the store and its personnel. The inspectors rate the stores on ten dimensions; perfect scores in all categories would lead to a score of 100 points. The dimensions include the store and personnel’s appearance, the demeanor of the personnel, the efficiency and effectiveness of the check-out system, among other things.

Note 2: NI means net income; # means number of; % means percentage.

REFERENCES BIBLIOGRAPHIQUES

Abelson, R.P., et A.Levi. (1985). Decision making and decision theory. In G.Lindsey and E.Aronson (Eds.), *The handbook of social psychology* (Vol.1, pp. 231-309).New York: Random House.

Banker, R. D., H. Chang, et M. Pizzini. (2004). « The Balanced Scorecard : Judgmental effects of performance measures linked to strategy ». *The Accounting Review* 79 (1) : 1-23.

Beach, L. R., et T. R. Mitchell. (1978). « A contingency model for the selection of decision strategies ». *Academy of Management Review* 3: 439-449.

Bettman, J. R. et P. Kakkar. (1977). « Effects of information presentation format on consumer information acquisition strategies ». *Journal of Consumer Research* 3: 233-240.

Cauvin, E., B. R. Neumann, et M. L. Roberts. (2007). « Financial measures bias in evaluating corporate managers' performance ». Working paper, University of Colorado at Denver and Health Sciences Center.

Dilla, W. N., et P. J. Steinbart. (2005). « Relative weighting of common and unique balanced scorecard measures by knowledgeable decision makers ». *Behavioral Research in Accounting* 17 : 43-53.

Halford, G. S., R. Baker, J. E. McGredden, et J. D. Bain. (2005). « How many variables can humans process ? » *Psychological Science* 16 (1) : 70-76.

Ittner, C. D., D. F. Larcker, et M. W. Meyer. (2003). « Subjectivity and the weighting of performance measures : Evidence from a Balanced Scorecard ». *The Accounting Review* 78 (Juillet) : 725-758.

Kaplan, R. et D. Norton. (1996). *The Balanced Scorecard*. Boston, MA: Harvard Business School Press.

Krumwiede, K. R., T. V. Eaton, et M. R. Swain. (2005). « A research note on the effects of target focus on the use of financial and nonfinancial measures in Balanced Scorecard evaluations ». Working paper, Boise State University.

Libby, T., S. Salterio, et A. Webb. (2004). « The Balanced Scorecard : The effects of assurance and process accountability on managerial judgment ». *The Accounting Review* 79 (3).

Lipe, M., et S. Salterio. (2000). « The balanced scorecard : Judgmental effects of common and unique performance measures ». *The Accounting Review* 75 (3) : 283-298.

Lipe, M. G., et S. Salterio. (2002). « A note on the judgmental effects of the balanced scorecard's information organization ». *Accounting, Organizations and Society* 27 (6) : 531-540.

Luft, J., et M. Shields. (2001). « The effects of financial and nonfinancial performance measures on judgment and decision performance ». Working paper, Michigan State University.

Malina, M. A., et F. H. Selto. (2001). « Communicating and controlling strategy : An empirical study of the effectiveness of the Balanced Scorecard ». *Journal of Management Accounting Research* 13 : 47-90.

March, J. G. (1978). « Bounded rationality, ambiguity, and the engineering of choice ». *Bell Journal of Economics* 9: 587-608.

Payne, J. W., J. R. Bettman, et E. J. Johnson. (1988). « Adaptive strategy selection in decision making ». *Journal of Experimental Psychology : Learning, Memory, and Cognition* 14 (3) : 534-552.

Roberts, M. L., T. A. Albright, et A. R. Hibbets. (2004). « Debiasing the Balanced Scorecard ». *Behavioral Research in Accounting* 16 : 75-88.

Russo, J. E., et B. A. Doshier. (1983). « Strategies for multiattribute binary choice ». *Journal of Experimental Psychology : Learning, Memory, and Cognition* 9 : 676-696.

Schiff, A. D., et L. R. Hoffman. (1996). « An exploration of the use of financial and nonfinancial measures of performance by executives in a service firm ». *Behavioral Research in Accounting* 8 : 134-151.

Silk, S. (1998). « Automating the balanced scorecard ». *Management Accounting* (Mai) : 38-44.

Slovic, P. (1975). « Choice between equally valued alternatives ». *Journal of Experimental Psychology : Human Perception and Performance* 1 : 280-287.

Slovic, P., et D. MacPhillamy. (1974). « Dimensional commensurability and cue utilization in comparative judgment ». *Organizational Behavior and Human Performance* 11 : 172-194.

Todd P. A., and I. Benbasat. (1994). « The influence of decision aids on choice strategies under conditions of high cognitive load ». *IEEE Transactions on Systems, Man, and Cybernetics* 24 (4) : 537-547.

Tversky, A. (1969). « Intransitivity of preferences ». *Psychological Review* 76 : 31-48.