



Munich Personal RePEc Archive

Lesson from Meta-analysis for the relation between budgetary participation and managerial performance

Bonache, Adrien and Maurice, Jonathan and Georgescu,
Irène

CREGOR.Université montpellier II,ENS-CACHAN,
ERFI.Université montpellier I,ENS-CACHAN, ERFI.Université
montpellier I, CEROM.ESC Montpellier

21 January 2009

Online at <https://mpra.ub.uni-muenchen.de/12958/>
MPRA Paper No. 12958, posted 23 Jan 2009 18:47 UTC

ENSEIGNEMENTS D'UNE META-ANALYSE SUR LE LIEN PARTICIPATION BUDGETAIRE-PERFORMANCE MANAGERIALE

Résumé

Objectifs. Cette méta-analyse se veut être une comparaison critique de trois méthodologies de synthèse quantitative. Elle permettra de mettre en application, en contrôle de gestion, certaines préconisations du forum international de Potsdam (1994) sur la méta-analyse. **Conception de l'étude.** Nous présentons, après un bref état de l'art classique sur le lien « participation budgétaire-performance managériale », trois méthodologies de méta-analyse en présence d'articles fortement hétérogènes quant à leurs méthodes et terrains. Après application des méthodes des votes et de la combinaison des p-values, nous appliquons sur nos articles la « best-evidence synthesis » recommandée au forum de Potsdam. **Résultats.** Les résultats des méthodes de méta-analyse mises en œuvre ne permettent pas de conclure sans prudence. Le lien semble négatif pour les articles de meilleure qualité.

Abstract

Objectives. This meta-analysis is a critical comparison of three of quantitative synthesis. It applies some propositions of the Potsdam international consultation on meta-analysis, in the field of management control. **Study design.** After a traditional narrative review, three meta-analysis methodologies for the synthesis of articles with heterogeneous methodologies and fields are presented. After a short critical presentation of our results with “box score” and p-values combination methods, we show the result of best-evidence synthesis and its stability as recommended at the Potsdam international consultation on meta-analysis. **Results.** The results of the methods of meta-analysis implemented do not make it possible to conclude without prudence. The link “budget participation-managerial performance” seems negative for the articles of better quality.

Mots clés : méta-analyse, participation budgétaire, performance managériale, méthodes des votes, synthèse des meilleurs études

Key words: meta-analysis, budgetary participation, managerial performance, box score methods, best-evidence Synthesis

Bienfaits et origines de la participation.

La littérature en comptabilité et en management en général souligne l'importance de la participation dans les problématiques d'établissement des budgets. Dans cette littérature, on voit que la participation à l'élaboration du budget des managers, se trouvant en milieu ou en bas de la ligne hiérarchique, peut avoir deux effets positifs (Banovic, 2005).

Premièrement, grâce à la participation au budget, des managers du bas de ligne hiérarchique se sentent plus impliqués. Ainsi, ils sont plus à même de respecter les objectifs fixés (Welsh, Hilton et Gordon, 1988).

Deuxièmement, la participation permet de diminuer l'asymétrie d'information. Ce qui permet à ceux se trouvant en haut de la ligne hiérarchique d'améliorer leurs connaissances du terrain, qui sont en principe mieux maîtrisées par ceux se trouvant en bas de la hiérarchie (Sivaramakrishnan, 1994).

La problématique des antécédents de la participation fut traitée par Shields et Shields (1998). Pour eux, on fait participer les managers « pour planifier et établir des buts en cas d'environnements incertains, pour motiver les subordonnés lorsqu'il y a des incertitudes concernant les tâches à accomplir mais aussi pour coordonner les tâches en cas d'interdépendance entre celles ci »¹.

Somme toute, il semblerait possible d'affirmer, comme Argyris (1952, p.28), que la participation est le principal remède aux effets pervers associés au processus budgétaire. D'après ces travaux en contrôle de gestion, pourrait-on inférer qu'une participation effective des représentants des Etats membres à la détermination du budget communautaire améliorerait l'atteinte des objectifs fixés ?

Remise en question de la participation « panacée ».

Il semblerait que certains managers utilisent cette opportunité pour introduire des biais dans les nombres qu'ils fixent et transmettent. Aussi, ils seront particulièrement enclin à diminuer les standards d'objectifs et d'augmenter leur besoin en ressources ; même si, cette propension au « slack » semblerait être modérée par la pression sociale (Young, 1985, p. 830).

Par ailleurs, dans de nombreuses entreprises, on retrouve de la pseudo-participation plutôt que de la participation. C'est à dire que l'on implique les subordonnés sans leur laisser de pouvoir, au lieu de leur laisser, de surcroît, la possibilité d'influencer l'établissement du budget (Becker et Green, 1962 et 1964).

Il semblerait que la participation n'ait pas un impact aussi clair que cela sur la performance lorsqu'elle permet de biaiser les budgets ou de créer un effet « Hawthorne » en donnant une illusion de participation qui peut casser la confiance à long terme et donc avoir un impact négatif sur le résultat de l'entreprise. Aussi, sur le plan théorique, peut-être semblerait-il plus pertinent de s'appuyer sur une version du modèle principal-agent de Penno (1990) montrant que la participation peut ne pas avoir de valeur ? Au delà d'un intérêt théorique, cette remise en question du lien positif participation-performance dans l'élaboration des budgets pourrait-elle plaider en faveur d'une moindre participation des chefs d'Etats membres au budget européen ?

¹ « Participative budgeting exists for planning and goal setting when there is environmental uncertainty; it exists for motivating subordinates when there is task uncertainty; and, it exists for co-ordinating interdependence when there is task interdependence » (Shields et Shields, 1998, p.65). T.D.A.

Les problématiques autour de la participation.

Il existe deux types de travaux autour de la question de la participation au budget par les subordonnés. Les premiers sembleraient s'intéresser à trouver les conditions « optimales » ou contingentes de participation au budget. D'autres travaux ont essayé d'établir des liens entre participation et d'autres variables comme la performance, le « slack », la satisfaction... Face à cette diversité de travaux aux conclusions souvent divergentes, il semble aussi possible de soulever un intérêt méthodologique : peut-on synthétiser ces travaux pour éclairer le décideur qui veut choisir entre faire participer ou non toutes les parties prenantes au budget ?

Les consensus bien établis sur les conditions... friables sur la relation participation-performance.

Les premiers travaux semblent trouver un consensus autour de la structuration des organisations -plus elles sont structurées et grandes plus il y a besoin de participation- (Burns et Waterhouse, 1975) et de la nature inertielle de la participation au budget -il faut y avoir participé, pour désirer y participer- (Hofstede, 1968).

Le consensus semble plus fragile concernant le lien entre performance et participation. En effet, bien que rarement mis en doute avant les années 1980, ce lien n'a que rarement fait l'objet d'études empiriques. Et de ce fait, cette question ne reste aujourd'hui pas tout à fait tranchée. Brownell (1981) trouve, lors d'une expérience avec 46 étudiants, que ce lien dépend de la personnalité (« locus of control ») du manager : la participation accroîtrait la performance, seulement pour ceux pensant avoir leur destin en main. La validation externe de ce résultat fut établi puisqu'il corrobore son intuition en réalisant la même étude sur 48 managers de la baie de San Francisco (Brownell, 1982).

Vers une complexification de la représentation théorique de la relation performance-participation.

Suivant les travaux de Brownell (1981, 1982a et 1982b), certains auteurs cherchent une explication contingente de ce lien pour expliquer la non significativité des résultats. Nouri et Parker (1998) mettent en avant l'implication organisationnelle et l'adéquation du budget comme facteurs de contingence ; Frucot et Shearon (1991) et Tsui (2001) soulignent l'importance de facteurs culturels dans ce lien et enfin Wentzel (2002) pense que la perception de justice et l'implication dans les buts contingentent l'impact de la participation sur la performance.

Nonobstant, d'autres auteurs ne trouvant pas de relations significatives pensent que ce lien est bien plus complexe (Shields et Young, 1993). Kren (1992) et Chong et Chong (2002) montrent que la participation ne joue sur la performance qu'en impactant l'implication, qui elle-même influence l'information pertinente sur le travail (« job-relevant information ») et cette information affecte enfin positivement et significativement la performance.

Éléments de clarifications conceptuelles.

Notons que les mesures de la performance et de la participation divergent énormément suivant les travaux mais qu'une validité de consensus semble s'être établie autour de deux échelles de mesures et définitions. Aussi, suivant ce consensus, la performance sera entendue dans la suite de ce papier comme une performance managériale, mesurée avec l'échelle de Mahoney (1963) et la participation est entendue comme une participation budgétaire, mesurée avec l'échelle de Milani (1965). Cela expliquera que nous excluons de la méta-analyse des meilleures études,

celles ne retenant pas ces deux échelles et donc ces deux définitions étroites de la participation et de la performance.

Nous définissons la méta-analyse comme Olkin (1995) en reprenant la définition de la « National Library of Medicine » : c'est « une méthode quantitative combinant les résultats d'études indépendantes (en principe tirées de la littérature publiée) et synthétisant les résumés et conclusions qui pourraient servir à évaluer l'efficacité d'une [pratique], planifier des nouvelles études, etc, avec des applications principalement dans le champ de la recherche... » (p. 133). Cette forme de synthèse vise à palier les limites des revues de littérature classiques. Celles-ci sont en effet rarement exhaustives, elles sont souvent biaisées par la tendance consciente ou inconsciente de mettre en avant ses papiers ou ceux de ses collègues, on n'y décrit que rarement la procédure de recherche d'articles, procédure qui est rarement systématique (Slavin, 1995). Pour systématiser ces revues de littérature, beaucoup sont passés par la méthode des votes qui compte le nombre d'études dont les résultats sont positifs-significatifs, positifs-non significatifs, négatifs-significatifs et négatifs-non significatifs. Si les résultats positifs-significatifs dépassent les autres, on dira que le lien est effectif et positif. Cette méthode est inadéquate car elle donne trop de poids aux grandes études (plus on a de répondants, plus les résultats sont significatifs), elle ne prête pas attention à la qualité des articles et ne différencie pas les effets importants et faibles (Slavin, 1995).

La méta-analyse fait croire, en étant exhaustive, qu'elle échappe au biais de publication mais souvent, elle n'a que l'illusion d'une exhaustivité (on recense souvent les travaux français, anglais, espagnols voire allemands, alors qu'il existe des travaux dans d'autres langues). De plus, lorsqu'elle a comme unité d'analyse la mesure (comme le recommande Glass, 1976) et non l'étude, cela vient biaiser ses résultats en faveur des études utilisant plusieurs mesures. En outre, l'exhaustivité vient à inclure des études de terrains dont la sélection n'a pas été « randomisée » ; ainsi, certains proposent de limiter l'application de la méta-analyse aux expérimentations (qui n'ont cependant qu'une faible validité externe) (Shapiro, 1993) mais d'autres, pourtant sceptiques à l'égard de la méta-analyse, disent qu'il faudrait améliorer la conception des études de terrain pour pouvoir réaliser des méta-analyses à forte validité interne et externe (Feinstein, 1995, p. 77).

Une autre solution est de ne retenir que les meilleures études pour faire une synthèse, ce que Slavin (1995) appelle la « best-evidence synthesis ». Mais, dans chaque discipline, se pose alors le problème de la définition des règles pour ne retenir que les meilleures études.

Aussi, la question principale de notre papier est la suivante : *Peut-on faire ressortir par le truchement d'une méta-analyse une conclusion univoque sur la relation participation-performance ?*

L'étude du lien participation budgétaire-performance fera apparaître dans un premier temps la méthodologie et les articles utilisés. Puis, nous montrerons les résultats obtenus par le biais de plusieurs méthodologies et, enfin, nous discuterons de nos résultats et la pertinence des méthodes de méta-analyse.

1. METHODES DE META-ANALYSE ET ARTICLES ANALYSES

1.1. Articles étudiés et justification des méthodes de méta-analyse

Présentation des articles compilés.

Nous avons sélectionné les articles par le biais des revues de littérature faites par Shields et Shields (1998) et Banovic (2005). Nous avons ensuite complété ces revues par des recherches personnelles sur Scholar.google.com et sur JSTOR. Par ailleurs, un des auteurs du papier réalise un travail doctoral dans ce champ de recherche et nous a donc permis d'avoir de nombreux articles difficilement accessibles grâce à ses contacts.

Tableau 1 : Articles

Études	Échantillon	Mesures de la participation	Type*	Mesure de la performance	Corrélation	Significativité
Aranya, 1990	223 managers canadien	Echelle personnelle	I	Echelle personnelle	1	p<0,05
Brownell, 1981	Expérimentation : 46 étudiants, 48 managers USA		I		-1	p<0,01 (dépend du « locus of control »)
Brownell, 1982	48 managers USA	Hofstede (1967) et Milani (1975)	I	Mahoney (1963/1965)	1	p<0,01
Brownell, 1982	48 managers USA	Hofstede (1967) et Milani (1975)	M	Mahoney (1963/1965)	-1	p<0,01
Brownell, 1985	61 managers USA	Milani (1975)	I	Mahoney (1963) scale	1	n.s.
Brownell et Dunk, 1991	79 managers USA	Milani (1975)	M	Mahoney (1963) scale	-1 -1 -1 -1 1 1 1	p = 0,28 p = 0,374 p = 0,202 p = 0,004 p = 0,027 p = 0,165 p = 0,031
Brownell et Hirst, 1986	76 managers USA	Hofstede (1967) et Milani (1975)	M	Mahoney et al.	1	n.s
Brownell et Mc Innes, 1991	224 managers USA	Milani(1975)	I	Mahoney	1 1	p<0,05 (Hofstede) p<0,01(Milani)
Brownell et Merchant, 1990	146 gestionnaires de production USA	Echelles personnelles	I	Echelles personnelles	1	n.s.
Cherrington et Cherrington, 1973	Expérimentation : 230 étudiants	Echelle propre	I	Atteinte d'objectifs	1	p<0,0001
Chong et Chong, 2002	79 managers Australie	Milani (1975)	I	Mahoney (1963)	1	n.s.
Dunk, 1990	26 managers USA	Milani (1975)	I	Mahoney (1963)	1	p = 0,0087
Dunk, 1993	79 managers USA	Milani (1975)	M	Mahoney (1963)	1 -1	p = 0,009 p = 0,588
Frucot et Shearon, 1991	83 managers mexicains	Milani (1975)	I	Mahoney (1963)	-1	p<0,05
Gul, Tsui, Fong et Kwok, 1995	54 managers de Hong Kong	Milani(1975)	I	Mahoney (1963)	-1	p = 0,004
Kenis, 1979	169 managers USA	Echelle propre	I	Echelle propre	1	p<0,01
Kren, 1990	Expérimentation: 44 étudiants	Protocole expérimental propre	I	Protocole expérimental propre	1	p = 0,0039

Kren, 1992b	80 managers Fortune 500	Milani (1975)	I	Mahoney (1963)	1	p = 0,07
Lau, Low et Eggleton, 1995	112 managers de GE de Singapour	Milani (1975)	M	Mahoney (1963)	-1	p = 0,006
Merchant, 1981	19 entreprises électroniques USA	Trois échelles propres	I	Auto évaluation des managers par rapport à la moyenne	1 (influence) autres mesures -1 ou +1	p<0,05 autre mesures n.s.
Merchant, 1984	170 managers USA	Facteur : influence sur la planification budgétaire	I	Mesure subjective de la performance : échelle propre	1 1	p = 0,049 p = 0,099
Mia, 1988	83 managers USA	Milani (1975)	I	Notation des supérieurs, échelle de Likert	-1 -1	p = 0,43 p = 0,24
Mia, 1989	62 managers USA	Milani(1975) / Brownell (1979)		Notation des supérieurs, échelle de Likert	1	p = 0,11
Milani, 1975	82 contremaîtres	Propre échelle	I	1/Performance relative de l'unité 2/ « hour to base »	1	Dépend du mois et de la mesure de performance
Nouri et Parker, 1998	135 managers	Milani (1975)	I	Govindarajan et Gupta (1985)	1	p<0,001
Tiller, 1983	Expérimentation : 150 étudiants	Echelle de 30 points classant les réponses sur le sentiment de liberté dans la fixation du standard	I	Evaluation en unité monétaire propre à l'expérimentation	1	n.s. (p=0,149)
Tsui, 2001	89 managers en Chine de GE	Milani (1975)	I	Mahoney (1963)	1 globale-ment -1 pour chinois 1 pour occidentaux	p<0,001 ns
Wentzel, 2002	74 managers d'un grand hôpital USA	Milani (1975)	I	Mahoney (1963)	1	n.s.

*I = participation budgétaire indépendante, M = participation budgétaire modératrice

Choix et adaptation d'une méthodologie de synthèse.

Compte tenu de la diversité des échelles et des traitements statistiques que nous avons pu rencontrer au cours de notre revue de la littérature, il nous a semblé souhaitable de privilégier, dans un premier temps, une méthode de méta-analyse permettant d'inclure toutes les études empiriques disponibles sur le sujet, malgré certaines restrictions. Par exemple, les études de Milani (1975) et de Tiller (1983) ne faisaient pas apparaître la pente de la relation et son écart-type. Aussi, nous nous sommes retrouvés face à un arbitrage entre une perte d'articles et donc d'exhaustivité de notre méta-analyse et une restriction dans le choix de la méthode de méta-analyse. Finalement, nous avons préféré utiliser, dans un premier temps, la méthode des votes qui permettent de décider d'un effet global de cette relation, par delà les contingences et les aléas des expérimentations et malgré l'absence de certaines informations. Par ailleurs, nous avons combiné à cette méthode, différentes techniques de combinaison des degrés de significativité afin que le lecteur puisse se faire librement une idée du degré de significativité de nos résultats. Par prudence, nous avons fait jouer le degré d'inclusion et de non-inclusion des études en fonction de leur degré de significativité. Cela permettra de montrer les limites d'une méta-analyse qualitative de ce type dont le degré de sélection des études peut faire varier le résultat. Cela permettra de voir si comme le dit Slavin (1995, p. 10) ces méthodologies sont inadéquates.

Enfin, dans un second temps, nous laissons tomber l'impératif d'exhaustivité, et nous préférons n'inclure que les articles de qualité dans le cadre d'une synthèse des meilleures études en suivant les pistes pour l'élaboration d'un filtre de Slavin (1995) et de toutes les communications du forum de Potsdam (1994). Ce filtre prend en compte la pertinence des articles par rapport à leur cadre conceptuel, leur validité interne et externe et le nombre de répondants à l'étude.

1.2. Présentation des techniques de la méthode des votes et de combinaison des valeurs p

1.1.1. Calcul d'un effet total, de sa variance et de son intervalle de confiance

Codification des résultats

Associées à la méthode des votes, il existe deux modes de compilation des résultats des études.

Premièrement, on peut coder les résultats des articles sous forme ternaire : liaison significative positive (1), liaison significative et négative (-1) et liaison non significative (n.s.). L'avantage de cette méthodologie est qu'elle permet d'inclure dans la méta-analyse une barrière en fonction du degré de significativité. Ce qui permet d'éviter un biais introduit par des articles dont les relations ne sont pas bien établies, mais ça a l'inconvénient de rejeter les études bien conçues faites sur des petits échantillons.

Deuxièmement, on peut aussi coder les résultats en fonction du signe de la relation sans considération pour le degré de significativité. Cela présente l'avantage de ne pas exclure certaines études de la revue de la littérature. Mais le fait d'exclure ces études ne permet pas de montrer la faible significativité de notre résultat de synthèse puisque l'on s'assure *ex ante* de la significativité des études entrant dans notre analyse.

Pour dépasser ce dilemme, nous avons privilégié une codification permettant de faire entrer soit tous les résultats concernant le lien participation-performance, soit ne faire entrer dans notre synthèse de résultats que certains articles en faisant varier le degré de sélection η .

Le test des signes

Le test des signes permet de tester l'hypothèse de relation positive. Il s'agit de diviser le nombre de résultats positifs par le nombre total de résultats compris dans les articles étudiés. Si la relation entre les variables est positive, alors la probabilité d'obtenir un résultat positif est supérieure à 0,5. Le test des signes est donc une application du test non paramétrique binomial aux deux hypothèses :

$$H_0 : \pi = 0.5$$

$$H_1 : \pi \neq 0.5$$

Où π est la proportion d'essais positifs estimée à partir d'un nombre d'essais positifs u par $p = u/k$, k étant le nombre total d'essais.

Le présent test consiste à calculer, sous H_0 , la probabilité d'obtenir un nombre de coefficients positifs au moins égal au nombre observés. Cette probabilité est égale à la probabilité que la variable aléatoire « nombre de résultats positifs » U soit supérieure à la valeur observée u . U

suit donc une loi binomiale $B(1/2, k)$ sous H_0 .

Estimation d'un intervalle de confiance et variance de l'effet global sous la méthode des votes.

Le développement qui suit vaut pour des résultats dont l'effectif est similaire. C'est à dire, si n_i est l'effectif ayant permis d'atteindre le résultat de l'article i , alors formellement $\forall i, n_i = n$. Cette hypothèse est particulièrement restrictive. Nonobstant, Hedges et Olkin (1980) admettent que si les effectifs ne sont pas trop différents, il est possible de les considérer comme étant tous égaux à une « valeur moyenne ». Malgré que Hedges et Olkin proposent de retenir soit l'effectif moyen, soit l'effectif minimum, soit l'effectif maximum, nous préférons à leur solution celle de Gibbons, Olkin et Sobkel (1977) du carré de la moyenne des racines :

$$\bar{n} = \left(\frac{\sum_{i=1}^{i=k} \sqrt{n_i}}{k} \right)^2$$

La « valeur moyenne » de Gibbons et al. (1977) a l'avantage d'être moins influencée par les valeurs extrêmes que la moyenne arithmétique. Les effectifs allant de 26 à 230, il ne semble pas nécessaire d'appliquer les méthodes développées spécialement pour ces situations (où les échantillons sont très hétérogènes) dans l'ouvrage de Cooper et Hedges (1994).

Une fois ce problème de valeur moyenne dépassé, il s'agit d'estimer un effet global à partir des résultats sur la base d'un codage dichotomique. Soit T_i l'estimation inconnue de l'effet d'une variable sur une autre du i -ème résultat. Si T_i dépasse une variable critique K_α (par exemple $K_{5\%}$ pour un seuil de signification de 5%), le résultat de cet effet sera codé significatif-positif et significatif-négatif dans le cas contraire. Formellement, nous avons :

$$\begin{aligned} X_i &= 1 \text{ si } T_i > K_\alpha \\ X_i &= 0 \text{ si } T_i \leq K_\alpha \end{aligned}$$

On retombe alors sur notre propos au dessus et sur le test non paramétrique binomial :

$$\begin{aligned} \pi &= P(X_i = 1) = P(T_i > K_\alpha) \\ 1 - \pi &= P(X_i = 0) = P(T_i \leq K_\alpha) \end{aligned}$$

On utilise alors une approximation normale de la loi binomiale pour obtenir un intervalle de confiance autour de l'estimation p du nombre de résultats positifs dans un échantillon d'articles de taille k . La variance de p est donc estimée par :

$$V(p) = p \cdot \frac{1-p}{k}$$

L'intervalle de confiance, utilisant l'approximation normale, à $(1 - \alpha) \%$ est donc défini par :

$$\left[\underline{p} = \bar{p} \right] = \left[p - z_{\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{k}}; p + z_{\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{k}} \right]$$

où $z_{\alpha/2}$ représente le fractile d'ordre $\alpha/2$ de la loi normale centrée réduite.

1.1.2. Combinaison des degrés de significativité

P-value cumulée : interprétation et portée limitée.

La p-value pour chaque résultat, testant la relation d'une variable X indépendante/modératrice sur une variable Y dépendante, est le reflet du test d'une même hypothèse nulle, très générale : X n'est pas corrélé avec Y. C'est à dire que si l'on teste la relation $Y_i = a_i \cdot X_i + b_i$, pour un article i, on teste $H_0 : a_i = 0$.

Aussi, dans la combinaison des degrés de significativité, l'hypothèse nulle (dite « omnibus ») est :

$$H_0 : a_1 = a_2 = a_3 = \dots = a_k = 0$$

Un degré de signification cumulé significatif permet donc de rejeter l'hypothèse H_0 , c'est à dire que pour au moins un résultat, la pente de la droite de X en Y n'est pas nulle.

Aussi, l'intérêt du calcul du degré de signification cumulé est limitée mais garde l'avantage de pouvoir regrouper des études ayant envisagé et utilisé des types de critères divers et variés. Pour cumuler ces degrés de signification, on retrouve dans la littérature deux types de méthodes (Hedges et Olkin, 1985 ; Becker, 1994) :

- celles basées sur l'uniformité des distributions des p-values,
- celles basées sur des transformations des p-values.

Méthodes de combinaison sous hypothèse d'uniformité de la distribution des p-values.

Proposition : la distribution des degrés de signification sous l'hypothèse nulle est uniforme sur l'intervalle [0,1] (Fisher, 1932).

Sur la base de cette proposition, on peut utiliser deux techniques de combinaison des degrés de signification :

1) La méthode de la p-value minimum :

Hedges et Olkin (1985) ont mis en avant que pour tester H_0 , il semblerait possible d'utiliser une simple comparaison de la valeur minimale dans la distribution des p-values, p_{\min} , avec $1-(1-\alpha)^{1/k}$. On rejettera l'hypothèse « omnibus » si $p_{\min} < 1-(1-\alpha)^{1/k}$.

2) La méthode de la moyenne des p-values :

Hedges et Olkin (1985) montrent que la statistique suivante suit une loi normale centrée réduite :

$$V = \sqrt{12 \cdot k} \cdot \left(0.5 + \sum_{i=1}^{i=k} \frac{p_i}{k} \right) \longrightarrow N(0,1)$$

Ainsi, il suffit de lire la table de la loi normale centrée réduite pour connaître la valeur de la p-value cumulée et d'en déduire le degré de signification du résultat obtenu par la méthode des votes.

Méthodes de transformation des degrés de signification.

Là, il s'agit de transformer les p-values en valeurs de variables aléatoires en utilisant la fonction inverse de la distribution cumulative d'une loi (normale, de Fisher, de Student...). Si l'on note pv , une p-value particulière, l'opération consiste à transformer via l'inverse d'une loi F , pv en une variable aléatoire X . Formellement, nous avons : $X = F^{-1}(pv)$.

Il existe dans la littérature de nombreuses méthodes indirectes permettant de convertir des statistiques de tests diverses et variées en des valeurs aux caractéristiques identiques et de les combiner (Glass, 1976 ; Hunter et Schmidt, 1982). Nous avons décidé d'en utiliser deux très courantes :

1) La méthode de la somme des logarithmes :

Cette méthode s'appuie sur le fait que si une distribution v est uniformément distribué sur $[0,1]$, alors $-2 \cdot \ln(v)$ suit une loi du Chi 2 à deux degrés de liberté. En l'occurrence, nous utiliserons la statistique suivant un Chi 2 à $2k$ degré de liberté :

$$U = \sum_{i=1}^{i=k} -2 \cdot \ln(pv_i)$$

A partir de cette statistique, il suffit de lire la table du Chi 2 à $2k$ degrés de liberté pour obtenir la valeur du degré de signification cumulé. Si celle-ci est supérieure au risque de première espèce α , nous rejetterons l'hypothèse « omnibus » H_0 .

2) La méthode de la transformation normale :

Il s'agit de calculer la statistique Z qui suit une loi normale centrée réduite si H_0 est vraie :

$$Z = \frac{[\Phi^{-1} \cdot (pv_1) + \Phi^{-1} \cdot (pv_2) + \dots + \Phi^{-1} \cdot (pv_k)]}{\sqrt{k}} = \frac{\sum_{i=1}^{i=k} \Phi^{-1} \cdot (pv_i)}{\sqrt{k}}$$

L'hypothèse « omnibus » est donc rejetée si cette statistique excède la valeur appropriée de la loi normale standard, pour une erreur de première espèce donnée.

La réalisation de ces tests sur les articles traitant de la relation participation budgétaire - performance managériale permettra de mettre à l'épreuve des faits les propositions du forum de Potsdam qui admettent ces méthodes comme inadéquates parce qu'elles agrègent des résultats hétérogènes (Feinstein, 1995). De plus, elles combinent des résultats/mesures et non des études et retiennent comme filtre souvent le degré de significativité de chaque résultat. Elles donnent donc trop de poids aux études ayant un grand nombre de répondants (Slavin, 1995). Mais, nous allons quand même regarder ce que nous pouvons dire en suivant les recommandations du forum de Potsdam en appliquant la synthèse des meilleures études.

1.3. Filtre de qualité, mesure d'une amplitude d'effet et synthèse des meilleures études

Avant même d'appliquer une synthèse, il faut s'assurer que l'on ne fait pas de l'« alchimie » en

transformant du « rien en quelque chose », « du plomb en or » ou pire que l'on ne travaille pas avec des substances hétérogènes (Feinstein, 1995, p. 71). Feinstein met en garde contre l'application de la méta-analyse sur des observations mal conçues, non « randomisées » et hétérogènes. Au mieux, nous risquons de nous retrouver avec une « salade composée » en mélangeant des « pommes et des oranges », au pire avec des « substances moins savoureuses ».

Pour Feinstein, il faut respecter les principes scientifiques d'homogénéité sinon l'unité d'analyse disparaît ou au pire les résultats globaux qui se veulent d'une plus grande applicabilité sont une « mixture confuse » (p. 76). Pour lui, la solution n'est pas d'interdire la méta-analyse dans des champs de recherche où les études ne sont pas bien conçues, ni de faire de la sélection des meilleures études mais tout simplement de fixer des critères pour la réalisation de cas de meilleure qualité (Feinstein, 1995, p. 77).

Sa réflexion est intéressante et nous ne pouvons que conseiller à la communauté scientifique de suivre ces conseils. Dans notre cas où les études que nous avons récoltées sont de qualité diverse, une solution plus applicable est de sélectionner les études en fonction de leurs qualités et de ne retenir que celles qui semblent exceptionnelles en partant du principe que « cette sélectivité semble bien plus attractive que la combinaison des fruits, légumes ou autres choses dans une méta-analyse standard qui s'abstient même des précautions scientifiques offertes par la *randomisation* des cas sélectionnés » (Feinstein, 1995, p. 77). Aussi, un critère que nous avons retenu est celui de la randomisation. Toujours dans le soucis d'avoir une certaine validité interne, Slavin (1995, p. 13) conseille de juger la conception de l'étude pour s'assurer que les biais ont été minimisés. Aussi, nous avons lu les articles attentivement à deux et attribué une note en fonction du degré de *précision sur la conception du cas*. A savoir, qu'un article ne décrivant pas avec suffisamment de précision la constitution de son échantillon, sa collecte de données, le traitement fait à ces données se voit attribuer un score plus faible que l'article faisant cela avec soin. Retenons que ce critère est subjectif et qu'il doit au moins être donné, à bulletin secret, par deux individus pour diminuer tout biais individuel et collectif. Ce que nous avons fait pour cette méta-analyse, puis une discussion argumentée a suivie pour trouver une évaluation convenable.

Slavin souligne aussi que l'on doit s'assurer de la validité externe. Pour ce faire, il souligne qu'il est particulièrement important d'écartier toute expérimentation faite sur des étudiants en nombre très limité. Aussi, nous avons fixé deux critères pour s'assurer de la validité externe de notre étude. Premièrement, il faut que l'étude ne soit pas faite en laboratoire et sur des étudiants, si c'est le cas le score sur cet item sera de 0, sinon de 1. Deuxièmement, nous avons attribuée une note en fonction du nombre de personnes interrogées en nous centrant sur la médiane des publications ($n = 79$). Ce score est construit de la façon suivante, si les résultats de l'étude portait sur n_i personnes alors le score sera de $n_i / 79$.

Enfin et surtout, « le plus important principe d'inclusion doit être la pertinence par rapport au sujet d'étude » (Slavin, 1995, p.13). Aussi, il faut que les définitions retenues par l'étude soient exactement les mêmes que celles de la méta-analyse, que les mesures utilisées soient identiques et non pas proches ou reliées. Aussi, les études qui mesurent la performance managériale avec l'échelle de Mahoney obtiennent un score de 0.5, si elles mesurent la

performance avec une échelle s'inspirant de celle de Mahoney, 0,25 sinon 0. Puis, nous avons fait de même avec l'échelle de Milani comme référence, ce qui donne un « score de pertinence » sur un.

Tableau 2 : détermination du score de qualité*

Auteurs	Score de qualité	Random	Individus	Pertinence/échelle	Précision du design	Validité externe
Milani	3,5		1,0	0,5	1,0	1,0
Brownell Hirst	3,9		0,9	1,0	1,0	1,0
Brownell McInnes	4,4		1,4	1,0	1,0	1,0
Frucot Shearon	4		1,0	1,0	1,0	1,0
Kren	4,1	*	1,0	1,0	1,1	1,0
Shields Young	3,45		1,2	0,3	1,0	1,0
Nouri Parker	4,3		1,7	0,5	1,1	1,0
Tsui	4,1		1,1	1,0	1,0	1,0
Chong et Chong	4,3	*	1,0	1,0	1,3	1,0
Wentzel	4		0,9	1,0	1,1	1,0
Brownell (85)	3,7		0,7	1,0	1,0	1,0
Gull Tsui Fong Kwok	3,1		0,4	1,0	0,7	1,0
Lau Low Engelton	4,7	*	1,4	1,0	1,3	1,0
Mia (88)	3,8		1,0	0,5	1,0	1,0
Mia (89)	3,2		0,7	0,5	1,0	1,0
Dunk (90)	3,4	*	0,3	1,0	1,1	1,0
Dunk (93)	4,3	*	1,0	1,0	1,3	1,0

*Les critères furent fixés avant la lecture des articles en détail comme conseillé par Slavin (1995) pour éviter tout biais.

Puis suite à l'attribution de ce score de qualité, lorsque cela était possible compte tenu de la présentation des résultats, nous avons calculé des « tailles d'effet » (*effect size*). Ce calcul de taille d'effet fut impossible pour une étude de bonne qualité (Chong et Chong, 2002) et pour des études de moins bonne qualité (Milani, 1975 ; Shields et Young, 1993 ; Tsui, 2001). Pour calculer cette taille d'effet (θ_i pour l'étude i), il suffit de soustraire l'estimation de l'effet d'une variable sur l'autre que l'on trouve dans l'étude (le a de $Y = a \cdot X + b$) avec la valeur nulle (car sous H_0 : $a = 0$) puis de diviser cette différence avec l'écart type de l'estimation de l'effet a . Nous appliquons ce faisant la définition de Glass de cet effet « ...la combinaison des résultats des études [se fait] par le truchement du calcul d'une estimation de la taille de l'effet g qui est la différence standardisée entre la moyenne de l'expérience et celle du groupe de contrôle »² (Glass cité par Hedges et Olkin, 1985).

Enfin, nous calculons l'effet commun de la participation budgétaire sur la performance managériale :

$$\hat{\theta} = \frac{\sum_{i=1}^{i=n} q_i \cdot w_i \cdot \theta_i}{\sum_{i=1}^{i=n} q_i \cdot w_i}$$

² « ...combining the results of the studies by calculating an estimate of effect size g which is the standardized difference between the experimental and control group means. » T.D.A.

avec q_i le score de qualité défini au dessus, w_i l'inverse de la variance de l'estimation de l'effet dans chaque étude (les résultats statistiques ayant une forte variance sont faiblement pondérés).

Puis, on calcule sa variance :

$$V(\hat{\theta}) = \sum_{i=1}^{i=n} q_i^2 \cdot w_i / \left(\sum_{i=1}^{i=n} q_i \cdot w_i \right)^2$$

Ces deux statistiques permettent de faire un test d'association en calculant la statistique U égale au rapport du carré de l'effet commun $\hat{\theta}$ avec la variance de cet effet commun. Cette statistique suit un Chi 2 à un degré de liberté –ddl-, il suffit donc de la comparer au fractile de la loi du Chi 2 à 1 ddl. pour une erreur de première espèce de 5 % ou 1 % pour voir si l'effet commun de la participation budgétaire sur la performance managériale est significativement non nul. Si la statistique U est supérieur au fractile de la loi du Chi 2 à 1 ddl., alors l'hypothèse H_0 est rejetée et on peut conclure à un effet non nul significatif.

Si les meilleures études donne un effet commun différent de l'effet commun en présence de toutes les études, il faudrait conclure avec prudence nous précise Slavin (1995). Pour nous assurer de la robustesse de notre résultat, et surtout pour nous assurer que le score de qualité ne « biaise » pas la statistique, nous avons triangulé l'analyse en calculant la statistique U sans le score de qualité.

Dans les sections suivantes, nous présentons les résultats des différentes méthodologies de méta-analyse détaillées dans cette section, nous en discutons les portées et limites et nous tentons un rapprochement avec la gestion du budget communautaire.

2. RESULTATS, PORTEES ET LIMITES

2.1. Résultats d'une méta-analyse qualitative sur le lien participation-performance

Pour présenter brièvement le tableau 3, on aperçoit en haut de celui-ci un ensemble de seuils de signification des résultats. Nous avons fait nos calculs soit en faisant rentrer dans nos résultats toutes les études sur le sujet, soit seulement celles dont le lien entre participation et performance a une p-value inférieure à 10%, 5% et enfin 1%. Ceci permet de montrer que nous avons bien voulu rentrer toutes les études mais que nous avons pris en compte la qualité du lien en faisant varier le seuil de sélection des études. Cela permet de dépasser le dilemme entre pertinence des résultats et exhaustivité de la méta-analyse qualitative par la méthode des votes. Et cela permet de voir si la sélection sur le degré de significativité n'est pas un critère de qualité très pertinent en comparant les résultats de cette méthodologie avec ceux d'une synthèse des meilleures études.

Ensuite, nous avons défini avec la méthode des votes le nombre d'études montrant un lien positif entre participation et performance sur le nombre total d'études ayant traité ce lien. On s'aperçoit que lorsque l'on ne retient que les études ayant montré un lien très significatif, alors le nombre d'études montrant un lien positif devient plus faible, toujours supérieur à 0,5 mais

non significatif statistiquement. Pour les études ayant des résultats fortement significatifs sur le lien participation budgétaire-performance, on ne peut conclure à l'existence d'un lien significativement positif, contrairement aux cas où l'on retient des études ayant des résultats moins significatifs. Ce résultat ne fait que confirmer la portée pratique des méta-analyses qualitatives qui recèlent de nombreuses limites et biais que nous avons soulignés en première partie (Slavin, 1995).

Tableau 3 : Résultats de la méthode des votes et de combinaisons des p-values

Critère (η) :	Pas de seuil	$\alpha \leq 0,1$	$\alpha \leq 0,05$	$\alpha \leq 0,01$
Méthodes des votes				
Nombre d'études avec un lien positif	43	21	17	6
Nombre total d'études	61	28	23	11
Test du signe (p)	0,7	0,75	0,74	0,55
Probabilité d'obtenir la proportion p sous H0	0,00036	0,00186	0,00531	0,27441
Résultat du test du signe	rejet H0	rejet H0	rejet H0	acceptation H0
<i>Square mean root</i>	67,001	74,060	69,157	77,871
<i>Var(p)</i>	0,0034	0,0067	0,0084	0,0225
IC (borne inf.)	0,59	0,59	0,56	0,25
IC (borne sup.)	0,82	0,91	0,92	0,84
Combinaison des p-values				
<i>Méthode des p minimums (pmin)</i>	0,000039	0,000039	0,000039	0,000039
<i>Méthode des p min. (stat. de test)</i>	0,000841	0,001830	0,002228	0,004652
Méthode des p minimums (résultat)	rejet H0	rejet H0	rejet H0	rejet H0
<i>Méthode des p moyens (moyenne)</i>	0,1949	0,0294	0,0173	0,0037
<i>Méthode des p moyens (stat. de test)</i>	7,32	8,31	7,66	5,70
p cumulé	1,22E-013	0,00E+000	9,10E-015	5,93E-009
Méthode des p moyens (résultat)	rejet H0	rejet H0	rejet H0	rejet H0
<i>Méthode de la somme des log (somme)</i>	293,72	242,65	217,22	146,64
p combiné	6,56E-022	2,08E-026	1,78E-025	2,05E-020
Méthode de la somme des log (résultat)	rejet H0	rejet H0	rejet H0	rejet H0
<i>Méthode de la transformation normale (somme N¹)</i>	-65,59	-58,56	-51,51	-32,68
<i>Méthode de la transformation normale (stat. de test)</i>	-9,47	-11,49	-11,24	-9,85
p combiné	1,44E-021	7,84E-031	1,30E-029	3,36E-023
Méthode de la transformation normale (résultat)	rejet H0	rejet H0	rejet H0	rejet H0

Par ailleurs, nous avons calculé un effectif moyen pour ce résultat (*square mean root*). Puis, par rigueur, nous avons tenu à calculer une variance de notre proportion de liens positifs et un intervalle de confiance autour de celle-ci, pour une erreur de première espèce α de 5%.

On peut s'apercevoir en regardant les intervalles de confiance que notre résultat n'est pas univoque si l'on ne retient que des études avec un très fort degré de significativité ($p < 0,01$). En effet, la borne inférieure de l'intervalle de confiance, dans ce cas, est en dessous de 0,5 (0,25). En moyenne, le lien est positif, mais il ne l'est que dans 55% des cas et n'est pas significativement différent de 0,5 selon le test de signe effectué.

Enfin, compte tenu de cette faiblesse de notre résultat, nous avons calculé, via plusieurs méthodes, le degré de significativité combiné (p-value de notre résultat) car il semblait nécessaire de montrer que notre résultat n'était pas juste le résultat de la sélection d'une méthodologie particulière. Il apparaît paradoxalement que notre résultat est fortement significatif quelles que soient les études retenues dans l'analyse puisque toutes les méthodes

de calcul de la p-value combinée concorde sur le rejet de l'hypothèse nulle omnibus décrite dans la section précédente. Lorsque l'on retient les études ayant des résultats dont les p-values sont aussi supérieures à 1% (pas de seuil, $\alpha \leq 0,1$, $\alpha \leq 0,05$), les deux tests (méthode des votes et combinaison des degrés de signification) sont donc convergents et attestent d'un lien significatif entre la participation budgétaire et la performance. En revanche, lorsque l'on ne retient que les études ayant des résultats significatifs (p-value < 1%), les différentes méthodes de combinaison des degrés de signification indiquent toutes qu'au moins une des études montre un lien significatif alors que la méthode des votes ne permet pas de mettre à jour un lien positif statistiquement significatif entre la participation et la performance.

Ces divergences de résultat amènent deux réflexions. La première est que la combinaison des degrés de signification offre un critère trop peu discriminant car il ne s'intéresse qu'à l'existence d'au moins un effet significatif sur l'ensemble des études. Il ne permet donc pas de conclure à l'existence d'un lien global positif ou négatif entre une variable et une autre. L'avantage cependant de cette méthode, qui est la contrepartie de son inconvénient, est qu'elle permet d'inclure des études utilisant des critères d'analyse variés, ce qui est souvent le cas en sciences humaines et sociales. La seconde réflexion amenée par nos résultats concerne la méthodologie de la méta-analyse qualitative dans son ensemble. Au vu de nos résultats, il semble qu'un biais de sélection peut venir impacter fortement les conclusions d'une méta-analyse de ce type. En faisant varier l'échantillon d'articles retenus pour l'analyse en fonction d'un critère « objectif » (la p-value des résultats des études), nous arrivons à modifier les conclusions de l'analyse et au final, cela nous empêche de pouvoir réellement trancher la question théorique sous-jacente. Pour tenter de dépasser cette impasse méthodologique qui pourrait être due au critère de sélection retenu, nous avons mis en œuvre d'autres critères de qualité plus adéquats dans le cadre d'une synthèse des meilleures études (Slavin, 1995).

2.2. Les résultats d'une application des conclusions du forum de Potsdam (1994)

Nous avons mis en œuvre ici la « best-evidence synthesis » de Slavin (1995) telle que préconisée par le forum de Potsdam. Nous avons effectué nos tests de deux manières différentes.

Premièrement, nous avons sélectionnés, parmi l'ensemble des articles dont nous disposions (cf. tableau 1), ceux qui remplissaient l'ensemble des critères évoqués en partie 1. Chaque critère devait être rempli au moins partiellement sauf pour le critère « validité externe » qui est rempli entièrement par tous les articles compte tenu de notre définition de la validité externe. Nous avons donc retenu les articles dont le score de qualité est évoqué dans le tableau 2. Par ailleurs, nous avons également re-sélectionné les articles qui avaient utilisés la randomisation pour leur analyse. Nous considérons en effet, comme beaucoup d'auteurs, que cet élément renforce la validité et la robustesse des résultats obtenus. Par conséquent, nous avons également mené une analyse sur ces articles dont le nombre se monte à 5 (cf. tableau 2). Dans cette analyse, nous avons dû exclure le papier de Chong et Chong (2002), pourtant excellent sur le plan des critères de qualité retenus mais ne traitant que des liens indirects entre participation budgétaire et performance managériale.

Deuxièmement, dans un souci d'objectivité, nous avons refait les mêmes tests sans tenir

compte du score de qualité retenu en suivant la méthodologie suivante. L'hypothèse H_0 testée est toujours celle de la nullité de l'effet de la participation budgétaire sur la performance managériale ($H_0 : \hat{\theta} = 0$).

Sans le score de qualité, nous avons :

$$\hat{\theta} = \frac{\sum_{i=1}^{i=n} \theta_i \cdot w_i}{\sum_{i=1}^{i=n} w_i},$$

$$Var(\hat{\theta}) = 1 / \sum_{i=1}^{i=n} w_i$$

La statistique de test est ici : $U = \left(\sum_{i=1}^{i=n} \hat{\theta}_i \cdot w_i \right)^2 / \sum_{i=1}^{i=n} w_i \longrightarrow \chi^2(1)$

Le tableau 4 présente le résultat de la méta-analyse lorsque l'on prend en compte le score de qualité et le tableau 5 présente ces résultats sans prise en compte du score de qualité.

Tableau 4 : méta-analyse et synthèse de meilleures études avec score de qualité

	Toutes les études utilisables	Meilleures études
$\hat{\theta}$ (effet commun)	0.3098	-0.4888
Var ($\hat{\theta}$)	0.002110	0.004477
Statistique de test U	45.50	53.37
Fractile χ^2 pour $\alpha = 1\%$	6.63	6.63
Résultat du test	Effet positif significatif	Effet négatif significatif

Tableau 5 : méta-analyse et synthèse de meilleures études sans score de qualité

	Toutes les études utilisables	Meilleures études
$\hat{\theta}$ (effet commun)	0.473	-0.139
Var ($\hat{\theta}$)	0.002084	0.004389
Statistique de test U	107.17	4.3761
Fractile χ^2 pour $\alpha = 1\%$ *pour $\alpha = 5\%$	6.63	3.84*
Résultat du test	Effet positif significatif	Effet négatif significatif à 5%

Lorsque l'on retient l'ensemble des études utilisables remplissant les critères de notre « best-evidence synthesis », le lien entre la participation budgétaire et la performance managériale est positif et significatif (seuil de 1%), que l'on prenne en compte notre score de qualité ou non.

En revanche, lorsque l'on ne retient que les études dont l'échantillon de test est réalisé de manière aléatoire, on observe que l'effet commun est négatif et significatif (à 1% en tenant compte de notre score de qualité et à 5% sans le score).

Il y a donc une différence de conclusion importante au sein de cette méthode de méta-analyse, montrant l'importance du choix des critères de sélection de la qualité des articles. Lorsque l'on sélectionne les études de très bonnes qualités (dont l'échantillon est randomisé, qui ne sont pas réalisées sur des étudiants, qui retiennent des échelles de mesures reconnues et dont la conception de l'étude est suffisamment détaillée), le lien étudié est négatif. Alors que lorsque l'on retient toutes les études, l'effet commun est positif, que ce soit par la méthode des votes, la combinaison des p-value ou la « best-evidence synthesis » sans le critère sur la randomisation.

On ne peut donc trancher clairement lorsque le résultat est différent entre deux synthèses de plus ou moins bonne qualité. Il se pourrait que notre critère de qualité ne soit pas à même de clore le « débat ». Et l'on observe alors que la méta-analyse qui est utilisée en principe pour établir un résultat commun clair n'est pas forcément à même de conclure distinctement en faveur d'un sens de l'effet. Et cela semble valable que l'on utilise une méthode de combinaison jugée inadéquate (méthode des votes) ou que l'on utilise la méta-analyse seulement sur les études de qualité remarquable.

Nous remarquons par contre la sensibilité de la significativité de notre résultat au score de qualité. En effet, alors que notre résultat du tableau 4 semblait trancher en faveur d'un lien négatif lorsque l'on ne retient que les meilleures études. Quand on effectue le test d'association sans pondérer les tailles d'effet avec le score de qualité, on voit alors que nous obtenons un résultat significatif pour une erreur de première espèce plus élevée (5% au lieu de 1%). Cela nous invite à d'avantage de prudence concernant le résultat de notre méta-analyse mais soulève quand même un doute sur la prétendue positivité du lien entre participation budgétaire et performance managériale. Remarquons que la sélection des articles en fonction d'un score de qualité défini a priori, retenu après la lecture des actes du forum de Potsdam, ne donne pas le même résultat qu'une sélection sur le degré de significativité. Cela renforce les propos de Slavin (1995) qui juge la sélection sur le degré de significativité comme peu pertinente car elle donne trop de poids aux articles dont l'échantillon est très important.

Conclusion

Finalement, il se pourrait que notre méta-analyse confirme le scepticisme de nombreux auteurs à l'égard de ces méthodologies de synthèse numériques. Premièrement, si l'on prend tous les articles quelque soit leur qualité ou leur méthodologie, il se pourrait que l'on obtienne une « salade composée » ou une « mixture » qui ne nous permet en rien de trouver un effet global puisque les études de mauvaises qualités ou utilisant des mesures non conformes biaisent totalement le résultat global. Deuxièmement, lorsque l'on décide d'appliquer une combinaison sélective pour étudier l'effet global, le périmètre et les critères de sélections retenus pour combiner nos résultats ne sont pas aussi neutres que le prône Slavin (1995). Cela pourrait mettre en doute la scientificité de ce type de méthode et aller dans le sens de Feinstein (1995) qui faisait une allégorie en rapprochant la méta-analyse d'une alchimie transformant non pas le plomb en or, mais un ensemble de résultats non souhaitables en un résultat souhaitable. Peut être, serait-il plus souhaitable de suivre l'avis de Feinstein qui souligne que l'on ne pourra faire de la méta-analyse dans les sciences utilisant l'observation, comme le contrôle de gestion et la gestion, que lorsqu'on aura amélioré la qualité et le

contrôle des études de terrain réalisées. Il faudrait par exemple randomiser la sélection des terrains d'analyse, mesurer le phénomène avec des échelles de mesures faisant consensus et fiables et, entre autres, prendre des échantillons suffisamment importants.

Ainsi, les méta-analyses souvent utilisées par les hommes politiques pour avoir une synthèse des travaux de recherche avant de prendre une mesure politique ne semble pas à même de trancher le débat sur le lien entre participation budgétaire et performance managériale. Pourtant, on aurait pu voir un intérêt pour ces recherches dans l'avancement de la réflexion d'un réel budget communautaire pour l'Union Européenne. La structure du budget communautaire est actuellement remise en question sur le plan économique (Sapir et al., 2004) et la question de la participation budgétaire des Etats est à l'ordre du jour.

Compte tenu de nos conclusions, nous ne pouvons répondre que si l'on retient ce qui se fait de mieux en méta-analyse, c'est à dire en prenant en considération les pistes de réflexion du forum de Potsdam. Dans ce cas, le fait de faire participer des managers à l'élaboration d'un budget semble avoir un impact négatif sur la performance. Ce résultat doit cependant prendre en compte que beaucoup d'études de qualités diverses montrent le contraire et que d'éventuelles contingences globales pourraient ressortir d'une méta-régression sur le sujet. De plus, ce résultat dépend de nos critères de qualité. Il se pourrait que ce papier ne soit que le point de départ d'un programme de recherche visant à fixer des critères d'une étude de très bonne qualité en contrôle de gestion et en gestion en général. Une fois un consensus trouvé sur cette qualité et mis en œuvre dans des recherches sur le lien entre participation budgétaire et performance managériale, alors nous pourrions éventuellement répondre à notre question initiale.

Pour revenir sur la question du fédéralisme budgétaire au sein de l'Union Européenne, il faut toutefois souligner que le sens de ce lien dans les entreprises ne pourrait permettre d'inférer avec certitude le sens du lien entre participation au budget de l'Union et performance des Etats Membres dans l'atteinte d'objectifs communautaires. En effet, cette inférence théorique devrait ensuite être testée au niveau européen pour être éventuellement validée.

Références bibliographiques

- Aranya N. (1990) "Budget Instrumentality, Participation and Organizational Effectiveness", *Journal of Management Accounting Research*, 2(1):61-11
- Argyris C., (1952) "Graduate School of Business, Public Administration, and Cornell University", *The Impact of Budgets on People*. Controllership Foundation
- Banović D. (2005) "Evolution and critical evaluation of current budgeting practices", Master's thesis, Ljubljana University
- Becker S. et D. Green Jr. (1962) "Budgeting and Employee Behavior", *The Journal of Business*, 35(4):392-402
- Becker S.W. et D. Green Jr. (1964) "Budgeting and Employee Behavior: A Rejoinder to A Reply", *The Journal of Business*, 37(2):203-205
- Brownell P. (1981) "Participation in budgeting, locus of control and organizational effectiveness", *The Accounting Review*, 56(4):844-860
- Brownell P. (1982) "A field study examination of budgetary participation and locus of control", *The Accounting Review*, 57(4):766-777
- Brownell P. (1983) "Leadership style, budgetary participation and managerial behaviour", *Accounting, Organizations and Society*, 8(4):307-321
- Brownell P. (1985) "Budgetary systems and the control of functionally differentiated organizational activities", *Journal of Accounting Research*, 23(2):502-512
- Brownell P. et A.S. Dunk (1991) "Task uncertainty and its interaction with budgetary participation and budget emphasis: some methodological issues and empirical investigation", *Accounting, Organizations and Society*, 16(8):693-703
- Brownell P. et M. Hirst (1986) "Reliance on Accounting Information, Budgetary Participation, and Task Uncertainty: Tests of a Three-Way Interaction", *Journal of Accounting Research*, 24(2):241-249
- P. Brownell et M. McInnes (1986) "Budgetary Participation, Motivation, and Managerial Performance", *The Accounting Review*, 61(4):587-600
- P. Brownell et K.A. Merchant (1990) "The Budgetary and Performance Influences of Product Standardization and Manufacturing Process Automation", *Journal of Accounting Research*, 28(2):388-397
- Bruns W.J. et J.H. Waterhouse (1975) "Budgetary control and organization structure", *Journal of Accounting Research*, 13(2):177-203
- Chalos P. et S. Haka (1989) "Participative Budgeting and Managerial Performance*", *Decision Sciences*, 20(2):334-347
- Cherrington D.J. et J.O. Cherrington (1973) "Appropriate reinforcement contingencies in the budgeting process", *Journal of Accounting Research*, 11:225-253,.
- Chong V.K. and K.M. Chong (2002) "Budget Goal Commitment and Informational Effects of Budget Participation on Performance: A Structural Equation Modeling Approach", *Behavioral Research in Accounting*, 14(1):65-86
- Cooper H. et L. Hedges (1989) *The Handbook of Research Synthesis*. Russell Sage Foundation, The, 1994
- Dunk A.S. Budget emphasis, budgetary participation and managerial performance: a note. *Accounting, Organizations and Society*, 14(4):321-324,.

- Dunk A.S. (1990) "Budgetary Participation, Agreement on Evaluation Criteria and Managerial Performance: A Research Note", *Accounting, Organizations and Society*, 15(3):171-178
- Dunk A.S. (1993) "The Effect of Budget Emphasis and Information Asymmetry on the Relation Between Budgetary Participation and Slack", *Accounting Review*, 68:400-400
- Feinstein A.R. (1995) "Meta-analysis: Statistical alchemy for the 21st century", *Journal of clinical epidemiology*, 48(1):71-79
- Fisher R.A. (1932) *Statistical methods for research workers*
- Frucot V. et W.T. Shearon (1991) "Budgetary participation, locus of control, and Mexican managerial performance and job satisfaction", *The Accounting Review*, 66(1):80-99
- Gibbons J.D., I. Olkin, et M. Sobel (1977) *Selecting and ordering populations: A new statistical methodology*
- Glass G.V. (1976) "Primary, secondary, and meta-analysis of research", *Educational Researcher*, 5(10):3-8
- Glenn W.A., H. Ronald, et G. Paul (1988) "Budgeting: Prot Planning and Control", Prentice Hall International, 661:2001-2002
- Gul F.A., J.S.L. Tsui, S.C.C. Fong, et H.Y.L. Kwok (1995) "Decentralisation as a Moderating Factor in the Budgetary Participation-Performance Relationship: Some Hong Kong Evidence", *Accounting and Business Research*, pages 107-107
- Hedges L.V. et I. Olkin (1980) "Vote-counting methods in research synthesis", *Psychological Bulletin*, 88(2):359-369
- Hedges L.V. etI. Olkin (1985) *Statistical methods for meta-analysis*, Academic Press Orlando
- Hunter J.E., F.L. Schmidt, et G.B. Jackson (1982) *Meta-Analysis: Cumulating Research Findings Across Studies*, Sage Pubns
- Kang S.H., J. O'Brien, et K. Sivaramakrishnan. (1994) "Analysts' Interim Earnings Forecasts: Evidence on the Forecasting Process", *Journal of Accounting Research*, 32:103-103
- Kenis I. (1979) "Effects of budgetary goal characteristics on managerial attitudes and performance", *The Accounting Review*, 54(4):707-721
- Kren L. (1990) "Performance in a budget-based control system: an extended expectancy theory model approach.",*Journal of Management Accounting Research*, 2:100-112
- Kren L. (1992) "Budgetary participation and managerial performance: The impact of information and environmental volatility", *The Accounting Review*, 67(3):511-526
- Lau C.M., L.C. Low, et I.R. Eggleton. C. (1995) "The impact of reliance on accounting performance measures on job-related tension and managerial performance: additional evidence", *Accounting, Organizations and Society*, 20
- Lindquist T.M. (1995) "Fairness as an Antecedent to Participative Budgeting: Examining the Effects of Distributive Justice and Referent Cognitions on Satisfaction and Performance", *Journal of Management Accounting Research*, 7:122-47
- Merchant K.A. (1981) "The Design of the Corporate Budgeting System: Influences on Managerial Behavior and Performance", *The Accounting Review*, 56(4):813-829
- Merchant K.A. (1984) "Influences on Departmental Budgeting: An Empirical Examination of a Contingency Model", *Accounting, Organizations and Society*, 9(3/4):291-307

- Mia L. (1988) "Managerial attitude, motivation and the effectiveness of budget participation", *Accounting, Organizations and Society*, 13(5):465-475
- Mia L. (1989) "The impact of participation in budgeting and job difficulty on managerial performance and work motivation: a research note", *Accounting, Organizations and Society*, 14(4):347-357
- Milani K. (1975) "The Relationship of Participation in Budget-Setting to Industrial Supervisor Performance and Attitudes: A Field Study", *The Accounting Review*, 50(2):274-284
- Nouri H., G. Blau, et A. Shahid (1995) "The effect of socially desirable responding (SDR) on the relation between budgetary participation and self-reported job performance" *Advances in Management Accounting*, 4:163-77
- Nouri H. et RJ Parker (1998) "The relationship between budget participation and job performance: the roles of budget adequacy and organizational commitment", *Accounting, Organizations and Society*, 23(5):467-483
- Olkin I. (1995) "Statistical and theoretical considerations in meta-analysis", *Journal of Clinical Epidemiology*. 48(1):133-146
- Penno M. (1990) "Accounting Systems, Participation in Budgeting, and Performance Evaluation", *Accounting Review*, pages 303-314
- Sapir A., Ph. Aghion, G. Bertola, M. Hellwig, J. Pisani-Ferry, J. Viñals et H. Wallace (2004) « An Agenda for a Growing Europe, rapport au président de la commission européenne », Oxford University Press
- Shapiro S. (1993) "Meta-analysis/Shmeta-analysis. *American Journal of Epidemiology*", 140(9):771-778
- Shields J.F. et M.D. Shields (1998) "Antecedents of participative budgeting", *Accounting, Organizations and Society*, 23(1):49-76
- Shields J.F. et Young S.M. (1993) "Antecedents and consequences of participative budgeting: evidence on the effects of asymmetrical information", *Journal of Management Accounting Research*, 5(1), 265-280
- Slavin R.E. (1995) "Best evidence synthesis: An intelligent alternative to meta-analysis", *Journal of Clinical Epidemiology*. 48(1):9-18
- Tiller M. (1983) "The dissonance model of participative budgeting: an empirical exploration", *Journal of Accounting Research*, 21(2):581-595
- Tsui J.S.L. (2001) "The impact of culture on the relationship between budgetary participation, management accounting systems, and managerial performance An analysis of Chinese and Western managers", *International Journal of Accounting*, 36(2):125-146
- Wentzel K. (2002) "The Influence of Fairness Perceptions and Goal Commitment on Managers' Performance in a Budget Setting", *Behavioral Research in Accounting*, 14(1):247-271
- Yin R.K. (2003) *Case Study Research: Design and Methods*. Sage Publications Inc
- Young S.M. (1985) "Participative Budgeting: The Effects of Risk Aversion and Asymmetric Information on Budgetary Slack", *Journal of Accounting Research*, 23(2):829-842