

Enseignements d'une meta-analyse sur le lien participation budgétaire - performance managériale

Adrien Bonache, Jonathan Maurice, Karen Moris, Irène Georgescu

▶ **To cite this version:**

Adrien Bonache, Jonathan Maurice, Karen Moris, Irène Georgescu. Enseignements d'une meta-analyse sur le lien participation budgétaire - performance managériale. La place de la dimension européenne dans la Comptabilité Contrôle Audit, May 2009, Strasbourg, France. pp.CD ROM. halshs-00460130

HAL Id: halshs-00460130
https://shs.hal.science/halshs-00460130

Submitted on 26 Feb 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Adrien BONACHE
Allocataire Moniteur Normalien, *Université Montpellier II, Place Eugène Bataillon, 34 000 Montpellier, 04 99 92 05 23,* bonache@rip.ens-cachan.fr
Jonathan MAURICE
Allocataire Normalien - Moniteur, *Université Montpellier I, ISEM, Espace Richter, 34 960 Montpellier, 04 99 54 53 36,* jonathan.maurice@ens-cachan.org
Karen MORIS
ATER, *Université Lumière Lyon 2, FARGO-DIJON,* kmoris@yahoo.fr
Irène GEORGESCU
Assistant ESC Montpellier, *Université Montpellier I, ISEM, Espace Richter, 34 960 Montpellier, irenegeorgescu@neuf.fr*

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 est non-significatif 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 methods of meta-analysis implemented do not make it possible to conclude without prudence. The link "budget participation-managerial performance" is not statistically significant 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 management souligne généralement l'importance de la participation dans les problématiques d'établissement des budgets. Dans cette littérature, 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). Selon ces auteurs, la participation des managers est mise en œuvre « 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, la participation est le principal remède aux effets pervers associés au processus budgétaire (Argyris, 1952, p. 28).

Remise en question de la participation « panacée ».

Certains managers utiliseraient cependant cette opportunité pour introduire des biais dans les nombres qu'ils fixent et transmettent. Ainsi, ils seront particulièrement enclin à diminuer les standards d'objectifs et à augmenter leur besoin en ressources, même si cette propension au « slack » serait modérée par la pression sociale (Young, 1985, p. 830).

Par ailleurs, dans de nombreuses entreprises, la pseudo-participation est plus fréquente que la participation. Dans le cadre de l'élaboration des budgets, les subordonnés sont alors laissés sans réel pouvoir et n'ont pas la possibilité d'influencer l'établissement des budgets (Becker et Green, 1962 et 1964).

La participation n'aurait donc pas un impact aussi clair que cela sur la performance. Si la participation permet de biaiser les budgets ou de créer un effet « Hawthorne » (Roethlisberger et Dickson, 1939) en ne donnant qu'une illusion de participation, la confiance à long terme peut être affectée. Cela pourrait avoir un impact négatif sur le résultat de l'entreprise. Ainsi, sur le plan théorique, ne serait-il pas plus pertinent de s'appuyer sur une version du modèle principal-agent de Penno (1990) ? Ce modèle montre en effet que participation peut ne pas avoir de valeur. Si tel était le cas, cette absence de valeur impliquerait que les ressources engagées pour assurer la participation le sont inutilement ou pour un résultat faible. L'objet de cette étude va être de répondre globalement à la question de l'efficacité de la participation au niveau de la performance managériale. La réponse à cette question permettrait ainsi d'orienter le processus budgétaire vers plus ou moins de participation en fonction des objectifs poursuivis par les organisations.

Les problématiques autour de la participation.

Il existe deux types de travaux autour de la question de la participation au budget par les

[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.

subordonnés. Les premiers s'intéressent à trouver les conditions « optimales » ou contingentes de participation aux budgets. 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 y a aussi un intérêt méthodologique à notre papier : peut-on synthétiser les résultats 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 de 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).

Mais le consensus semble plus fragile en ce qui concerne 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. Lors d'une expérience avec 46 étudiants, Brownell (1981) trouve 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.

Eléments de clarifications conceptuelles.

Il faut noter 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. Ainsi, suivant ce consensus, la performance sera entendue dans la suite de ce papier comme une performance managériale, mesurée avec l'échelle de Mahoney (1963). La participation sera entendue comme une participation budgétaire, mesurée avec l'échelle de Milani (1965). Cela expliquera que nous exclurons de la méta-analyse des meilleures études, celles ne retenant pas au moins l'une de ces deux échelles.

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. Il n'y est que rarement décrit la procédure de recherche d'articles -procédure non systématique dans la plupart des cas (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 pourrait fait penser, en cherchant l'exhaustivité, qu'elle échappe au biais de publication[2]. Or, elle n'a que l'illusion d'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 : *les différentes méthodes de méta-analyse permettent-elles de conclure de façon univoque sur la relation participation budgétaire-performance managériale ?*

La section 1 présente les méthodologies de synthèse utilisées et les articles étudiés concernant le lien participation budgétaire – performance managériale. La section 2 présente et discute des résultats obtenus et de la pertinence des méthodes de méta-analyse.

1. METHODES DE META-ANALYSE ET ARTICLES ANALYSES

1.1. Littérature et justification des outils de synthèse

[2] « Le biais de publication vient du fait que les résultats statistiquement […] positifs et significatifs ont plus de chance d'être publiés et présentés à des congrès scientifiques que les résultats négatifs », Grégoire, Derderian et Lelaurier, 1995, p.159. T.D.A.

Présentation des papiers 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	Corré-lation	Significativité
Abdullah, 1998	62 managers Australie	Milani (1975)	I	Mahoney (1963)	1	p=0.01
Alam, Mia et Gnepa, WP	113 managers d'une ONG Bangladesh	Milani (1975)	I	Mahoney (1963)	1	p=0.02
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)	1	p<0,01
Brownell, 1982	48 managers USA	Hofstede (1967) et Milani (1975)	M	Mahoney (1963)	-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
Heath et Brown, 2007	171 employés de l'Oklahoma	Milani (1975)	I	Fraser (1995)	1	p=0.298
Jermias et Setiawan, 2008	204 managers d'une administration indonésienne	Vroom et Mann (1960)	I	Mahoney (1963)	-1 -1	p=0.04 p=0.13

Kenis, 1979	169 managers USA	Echelle propre	I	Echelle propre	1	p<0,01
Kobori, 2006	86 managers de Toyo Kenzai (Taïwan)	Hofstede (1967)	I	Mahoney (1963)	1	p=0.16
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 et Lim, 2002	83 managers australiens	Milani (1975)	I	Mahoney (1963)	1	p=0.16
Lau, Low et Eggleton, 1995	112 managers de GE de Singapour	Milani (1975)	M	Mahoney (1963)	-1	p = 0,006
Leach-Lopez, Stammerjohan et McNair, 2007	143 managers américains frontaliers et mexicains	Milani (1975)	I	Mahoney (1963)	1 1	p=0.0158 p=0.1657
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
Ni et Su, 2001	205 managers taïwanais	Milani (1975)	I	Mahoney (1963)	1 -1 1	p=0.96 p=0.51 p=0.27
Ni, Su, Chung et Cheng, 2005	155 managers de sociétés cotées taïwanaises	Milani (1975)	I	Mahoney (1963)	1 1 1	p=0.617 p=0.004 p=0.465
Nouri et Parker, 1998	135 managers	Milani (1975)	I	Govindarajan et Gupta (1985)	1	p<0,001
Parker et Kyj, 2006	70 managers banques et assurances	Milani (1975)	I	Mahoney (1963)	1	p=0.203
Su et Lin, WP	168 managers taïwanais	Milani (1975)	I	Mahoney (1963)	1	p=0.01
Subramaniam et Ashkanasy, 2001	114 managers australiens dans l'agroalimentaire	Milnai (1975)	I	Mahoney (1963)	-1	p=0.01
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 en première analyse la méthode des votes qui permet 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 associé 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 varier le critère 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 restreignons l'impératif d'exhaustivité en préférant n'inclure que les articles de qualité dans le cadre d'une synthèse des meilleures études. Nous suivons pour cela les pistes de Slavin (1995), et de toutes les communications du forum de Potsdam (1994), pour l'élaboration d'un filtre des articles. 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. Nous présentons l'ensemble des méthodes évoquées dans la sous-section suivante.

1.2. Présentation des techniques de la méthode des votes et de combinaison des « p-values »

1.2.1. 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. Cela permet d'éviter un biais introduit par des articles dont les relations ne sont pas bien établies, mais 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 :

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

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 :

$$X_i = 1 \text{ si } T_i > K_\alpha$$
$$X_i = 0 \text{ si } T_i \leq K_\alpha$$

Nous aboutissons alors au test non paramétrique binomial expliqué ci-dessus, exprimé de la façon suivante :

$$\pi = P(X_i = 1) = P(T_i > K_\alpha)$$

$$1 - \pi = P(X_i = 0) = P(T_i \leq K_\alpha)$$

Ensuite, une approximation normale de la loi binomiale est utilisée 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 – α) % est alors défini par :

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

où $z_{\alpha/2}$ représente le fractile d'ordre α/2 de la loi normale centrée réduite.
Dans une volonté de trianguler pour améliorer la fiabilité de nos conclusions, nous présentons au delà de cette méthode des votes un ensemble de méthodes permettant de combiner les degré de significativité (« p-values »). L'intérêt de présenter nos résultats avec l'ensemble des méthodes de combinaison des « p-values » est important puisque lorsque l'on infirme une croyance ou une hypothèse répandue, il faut s'assurer que cela n'est pas dû à la « lunette » ou technique utilisée pour ce faire.

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. Soit la relation $Y_i = a_i \cdot X_i + b_i$ pour un article i, nous testons ici H_0 : $a_i = 0$.

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

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

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.
L'intérêt du calcul du degré de signification cumulé est alors réduit, mais il garde l'avantage de pouvoir regrouper des études envisageant et utilisant des types de critères divers et variés.
Pour cumuler ces degrés de signification, la littérature recense 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, deux techniques de combinaison des degrés de signification

sont utilisables :

1) La méthode de la « p-value » minimum :

Hedges et Olkin (1985) ont mis en avant que pour tester H_0, il est 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 $p_{min} < 1-(1-\alpha)^{1/k}$.

Cette technique de test extrêmement simple consiste à regarder si la « p-value » minimum est significativement inférieure à l'erreur de première espèce alpha. Mais cette procédure ne s'intéresse qu'à une caractéristique de forme minimale. Il semble préférable de combiner à cette approche un test visant à vérifier l'uniformité de la distribution des « p-values », qui est une proposition de départ à cette méthode.

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.

Combiner ces deux techniques permet de s'assurer que la distribution des « p-values » est uniforme mais aussi que le minimum de cette distribution est significativement différent d'une erreur de première espèce fixée *ex ante*. Toutefois, il peut s'avérer intéressant de combiner à ce type de technique fondée sur l'hypothèse d'une uniformité de distribution, d'autres techniques se passant d'une telle hypothèse. Nous avons retenu deux méthodes de transformation de la distribution des « p-values » pour s'assurer que cela ne vient pas du type de transformation retenu. Lorsque l'on infirme une hypothèse répendue dans la littérature, tout comme lorsque l'on émet une nouvelle proposition, il faut s'assurer par une triangulation des méthodes que cette conclusion n'est pas issue d'un « montage statistique ».

Méthodes de transformation des degrés de signification.

Il s'agit ici 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 effet, Fisher (1932) écrit que l'on peut admettre qu'une « p-value » est un nombre aléatoire ayant autant de chance de sortir en tout point d'un intervalle [0,1]. L'ensemble des « p-values » approxime donc une fonction cumulative quelconque qui peut être la fonction de répartition d'une chi deux. En

outre, on peut approximer une inversion de la loi du chi deux en utilisant la transformation -2ln(.). Somme toute, nous utiliserons la statistique suivant un Chi 2 à 2k degré de liberté :

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

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.
Pour augmenter la fiabilité de nos résultats en s'assurant qu'ils ne viennent pas d'un certain type de transformation, il semble nécessaire d'appliquer une autre transformation fréquemment utilisée.

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 :

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

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 souvent comme filtre le degré de significativité de chaque résultat. Elles donnent donc trop de poids aux études ayant un grand nombre de répondants (Slavin, 1995). Sur la base de ces critiques, nous allons alors mettre en œuvre une méthode de synthèse normalement plus pertinente suivant les recommandations du forum de Potsdam : la synthèse des meilleures études. Cette méthodologie est justifiée et détaillée dans la soussection suivante.

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

Avant même de réaliser une synthèse, il faut s'assurer de ne pas faire 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. Dans le cas contraire, 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 de proscrire 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 analysées sont de qualité diverse, une solution plus applicable est de sélectionner les études en fonction de leur valeur. Il s'agira alors 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). Ainsi, nous avons retenu le critère 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. Nous avons alors 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, une discussion argumentée ayant suivi 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'écarter toute expérimentation faite sur des étudiants en nombre très limité. Ainsi, nous avons fixé deux critères pour s'assurer de la validité externe de notre étude. Premièrement, il faut que l'étude ne soit pas réalisée 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 = 81). 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 /81.

Enfin et surtout, « le plus important principe d'inclusion doit être la pertinence par rapport au sujet d'étude » (Slavin, 1995, p.13). Il faut donc 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. 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.

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 »[3] (Glass cité par Hedges et Olkin, 1985).

Tableau 2 : détermination du score de qualité*

Auteurs	Score de qualité	Random	Individus	Pertinence/échelles	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	**4,0**		1,0	1,0	1,0	1,0
Kren	**4,1**	*	1,0	1,0	1,1	1,0
Shields Young	**3,5**		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**		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)	**3,5**		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
Abdullah (98)	**4,0**	*	1,0	1,0	1,0	1,0
Kobori	**3,7**	*	1,1	0,5	1,1	1,0
Lau Lim	**4,6**		1,0	1,0	1,6	1,0
Ni Su	**5,3**	*	2,5	1,0	0,8	1,0
Leach-Lopez et al.	**4,2**		1,2	1,0	1,0	1,0
Leach-Lopez et al.	**3,6**		0,6	1,0	1,0	1,0
Alam Mia Gnepa	**4,4**		1,4	1,0	1,0	1,0
Ni et al.	**5,0**	*	1,9	1,0	1,1	1,0
Jermias Setiawan	**5,3**		2,5	0,5	1,3	1,0
Parker Kyj	**4,1**		0,9	1,0	1,2	1,0
Subramanian Ashkanasy	**4,6**	*	1,4	1,0	1,2	1,0
Su Lin	**5,3**	*	2,1	1,0	1,2	1,0
Heath Brown	**4,8**		2,1	0,5	1,2	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.

[3] « ...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.

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

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

avec q_i le score de qualité défini ci-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 :

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

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 supérieure au fractile de la loi du Chi 2 à 1 ddl., alors l'hypothèse H_0 est rejetée et nous pouvons conclure à un effet non nul significatif.

Si les meilleures études donnent un effet commun différent de l'effet commun en présence de toutes les études, Slavin (1995) nous précise qu'il faut conclure avec prudence. 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 précédemment, puis nous en discutons les portées et limites.

2. RESULTATS, PORTEES ET LIMITES

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

Le tableau 3 résume et classe nos résultats en fonction du seuil de sélection η reposant sur le degré de signification α. Nous avons obtenu ces résultats en faisant rentrer dans nos calculs soit 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%. Nous avons donc tenter de rentrer toutes les études tout en prenant 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 est un critère de qualité 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. 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 statistiquement significatif. 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 limitée des méta-analyses qualitatives, souffrant des biais que nous avons soulignés en première partie (Slavin, 1995).

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%.
En regardant les intervalles de confiance, 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 (0.39). En moyenne, le lien est positif, mais il ne l'est que dans 63% des cas et n'est pas significativement différent de 0,5 selon le test de signe effectué.

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

Critère (η) :	**Pas de seuil**	**α ≤ 0,1**	**α ≤ 0,05**	**α ≤ 0,01**
Méthodes des votes				
Nombre d'études avec un lien positif	57	26	22	10
Nombre total d'études	79	36	31	16
Test du signe (p)	0,72	0,72	0,71	0,63
Probabilité d'obtenir la proportion p sous H_0	0,00002	0,00197	0,00534	0,10506
Résultat du test du signe	rejet H_0	rejet H_0	rejet H_0	acceptation H_0
Square mean root	74,86	83,21	80,82	66,13
Var(p)	0,0025	0,0056	0,0066	0,0146
IC (borne inf.)	0,62	0,58	0,55	0,39
IC (borne sup.)	0,82	0,87	0,87	0,86
Combinaison des « p-values »				
Méthode des p minimums (pmin)	0,000039	0,000039	0,000039	0,000039
Méthode des p min. (stat. de test)	0,000649	0,001424	0,001653	0,003201
Méthode des p minimums (résultat)	rejet H_0	rejet H_0	rejet H_0	rejet H_0
Méthode des p moyens (moyenne)	0,2029	0,0263	0,0166	0,0039950
Méthode des p moyens (stat. de test)	9,15	9,85	9,32	6,87
p cumulé	0,00E+000	0,00E+000	0,00E+000	3,17E-012
Méthode des p moyens (résultat)	rejet H_0	rejet H_0	rejet H_0	rejet H_0
Méthode de la somme des log (somme)	383,14	304,80	279,37	167,61
p combiné	9,87E-021	2,07E-030	2,34E-029	2,64E-020
Méthode de la somme des log (résultat)	rejet H_0	rejet H_0	rejet H_0	rejet H_0
Méthode de la transformation normale (somme N^{-1})	-85,46	-74,33	-67,27	-37,78
Méthode de la transformation normale (stat. de test)	-9,61	-12,39	-12,08	-9,45
p combiné	0,00E+000	0,00E+000	0,00E+000	0,00E+000
Méthode de la transformation normale (résultat)	rejet H_0	rejet H_0	rejet H_0	rejet H_0

Enfin, compte tenu de cette faiblesse de notre résultat, nous avons calculé, via plusieurs méthodes, le degré de significativité combiné (« p-value » cumulée) 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 très 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. La méthode des votes ne permet pas de mettre à jour quant à elle un lien positif statistiquement significatif entre la participation et la performance.

Ces divergences de résultats 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), les conclusions de l'analyse sont modifiées. Au final, avec ces différentes méthodes qualitatives, nous ne pouvons 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 Cette méthodologie a été notamment présentée lors du forum de Potsdam (1994) dont nous faisons une application dans la sous-section suivante.

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 (Kendall et al., 1987 ; Feinstein, 1995), 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 11 (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 :

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

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

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

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)**	0.5408	-0.1029
Var ($\hat{\theta}$**)**	0.001765	0.003711
Statistique de test U	165.74	2.852
Fractile χ^2 pour α = 1%	6.63	6.63
Résultat du test	Effet positif significatif	Effet négatif non 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	0.1436
Var ($\hat{\theta}$**)**	0.002084	0.003621
Statistique de test U	107.17	5.6984
Fractile χ^2 pour α = 1%	6.63	6.63
Résultat du test	Effet positif significatif	Effet positif non significatif

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 non-significatif mais tantôt positif, tantôt négatif si l'on tient compte ou non de notre score de qualité.

Il y a donc une différence de conclusion 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 bonne qualité (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 non-significatif, positif sans score de qualité et négatif avec notre score de qualité. 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-values » ou la « best-evidence synthesis » sans le critère sur la « randomisation ».

On ne peut donc conclure sans prudence sur la positivité et la significativité du lien entre participation budgétaire et performance managériale. L’effet est positif et significatif si l’on retient l’ensemble des études disponibles, mais ce résultat reste discutable compte tenu des critiques à l’encontre des méthodes de méta-analyse qualitatives et quantitatives ne prenant pas en compte la qualité des papiers retenus. En revanche, le lien entre participation budgétaire et performance managériale n’est pas significatif lorsque l’on retient les meilleurs papiers ayant utilisé de la « randomisation ». L’impact de notre score de qualité est intéressant car nous voyons que celui-ci change le signe de l’effet qui devient négatif même s’il ne reste pas significatif, même à 5%. La non-significativité de l’effet est donc un résultat probant, qui nous permet de conclure que faire participer les managers au processus budgétaire ne semble pas avoir d’impact sur la performance managériale, dans le cadre des meilleures études réalisées sur le sujet. Toutefois, il convient d’attirer l’attention sur le caractère contingent de certains résultats des études synthétisées. Des facteurs culturels sont probablement à l’œuvre si l’on regarde l’origine géographique des populations étudiées (Asie, Etats-Unis, Europe).

Pour autant, des réflexions complémentaires sur les méthodologies de synthèse utilisées sont à mener. Tout d’abord, les résultats de la méthode des votes et de combinaison des « p-values » semblent trop peu discriminants. 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. Par ailleurs, la méthodologie de synthèse des meilleures études semble être une approche plus pertinente pour répondre à la question de recherche posée. Dans un premier temps, la systématisation de la recherche d'articles associée à l'évaluation de la qualité des études menées dans les articles sélectionnés permet de rétablir l'esprit critique du chercheur dans l'analyse. Cet esprit critique ne peut toutefois s'exercer que par la définition de critères de qualité objectifs et de les appliquer rigoureusement et systématiquement. Dans notre cas, lorsque l'on ne prend en compte que les études d'une forte validité externe et dont les mesures reposent sur des échelles comparables, nous trouvons que le lien participation-performance est positif et significatif. Il est possible toutefois que ce résultat souffre d'un biais de publication ou de méthode. En effet, ces études sont de bonne qualité mais ne sont pas effectuées sur des échantillons aléatoires, ce qui peut expliquer la positivité et significativité du lien que nous ne retrouvons pas lorsque nous durcissons les critères de qualité (en ajoutant la « randomisation » de l'échantillon notamment). Ensuite, lorsque ne sont retenues que les études ayant des méthodologies extrêmement proches et « randomisées », le lien devient non significatif. Le changement de signe du lien est toutefois intéressant à relever sur le plan de la méthode en montrant bien l'impact et l'intérêt de réaliser un score de qualité pour les articles (si ce score est bien conçu, alors il permet encore de mieux discriminer les articles mais s'il est conçu de façon approximative alors il est possible que le résultat de la méta-analyse soit modifié dans un sens non souhaitable).. Celui-ci doit donc reposer sur des critères vérifiables et objectifs pour ne pas retomber dans les travers des revues de littérature classiques, ce que nous avons montrer dans cette étude.

Conclusion

Compte tenu de notre analyse, il apparaît que les entrepreneurs ne peuvent pas faire participer dans l'espoir d'avoir un surcroit de performance. Cependant, il existe des contingences qui pourraient leur être favorables. Il pourrait être intéressant de mener une expérimentation en entreprise en faisant participer certains managers et pas d'autres. Des différences éventuelles de performance pourrait alors être décelées. Cela donnerait donc une information directement opérationnelle aux dirigeants de telle entreprise ayant regardé si le lien entre participation et performance était effectif et positif.

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 (avant d'ouvrir un marché à la concurrence par exemple) et les médecins (avant de prescrire un traitement) pour avoir une synthèse des travaux de recherche avant de prendre une décision ne semble pas à même de trancher le débat sur le lien entre participation budgétaire et performance managériale. On peut donc s'interroger sur l'utilité de la méta-analyse en sciences de gestion et surtout pour les managers et dirigeants. Cette question pourrait aussi être traitée en examinant si les professeurs des étudiants en gestion et les managers professionnels lisent plus ces travaux de synthèse que des articles contradictoires. Si ce n'est pas le cas, nous pourrions interroger l'utilité des synthèses en pratique, au delà du cercle restreint des universitaires.

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 non-significatif 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 oeuvre dans des recherches sur le lien entre participation budgétaire et performance managériale, alors nous pourrons éventuellement répondre à notre question initiale.

Références bibliographiques

Abdullah, S.N. (1998), “The role of motivation in the link between budgetary participation and managerial “, Analisis, 5(1&2):1-17

Alam B., Mia L. et Gnepa T.J., “Need for achievement, style of budgeting and managerial performance in a non government organization (NGO): evidence from an oriental culture”, working paper

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

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. et K.M. Chong (2002) “Budget Goal Commitment and Informational Effects of Budget Participation on Performance: A Structural Equation Modelling 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

Grégoire G., F. Derderian et J. Lelaurier (1995) “Selecting the language of the publications included in a meta-analysis: is there a tower of Babel bias”, Journal of Clinical Epidemiology, 48(1):159-163

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

Heath R.S. et J.F. Brown (2007), “A Re-Examination Of The Effect Of Job-Relevant Information On The Budgetary Participation-Job Performance Relation During An Age Of Employee Empowerment”, Journal of Applied Business Research, 23(1): 111

Hedges L.V. et I. Olkin (1980) “Vote-counting methods in research synthesis”, Psychological Bulletin, 88(2):359-369

Hedges L.V. et I. 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 Pub

Jermias J. et Setiawan T. (2008) “The moderating effects of hierarchy and control systems on the relationship between budgetary participation and performance”, International Journal of Accounting, 43(3):268--292

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

Kendall M.G., A. Stuart et J. K. Ord (1987) Kendall’s Advanced Theory of Statistic. London Charles Griffen.

Kobori K. (2006), “The Effect of Budgetary Participation on Performance: Role Ambiguity and Organizational Commitment as an Intervening Variable”, National Cheng Kung University Department of Business Administration Master Thesis, juillet 2006 (Titre original: “預算參與對管理績效間的影響之研究—以角色模糊和組織承諾為干擾變数”)

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. et E.W. Lim (2002) “The effects of procedural justice and evaluative styles on the relationship between budgetary participation and performance”, Advances in Accounting, incorporating Advances in International Accounting, 19:139-160

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

Leach-Lopez M.A., W.W. Stammerjohan et F.M. McNair (2007) “Differences in the Role of Job-Relevant Information in the Budget Participation-Performance Relationship among US and Mexican Managers: A Question of Culture or Communication”, Journal of Management Accounting Research,19(1):105-136

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

倪豐裕，蘇志泰 (2001) “組織規模和動機對於參與式預算制度效能的影響”, Commerce & Management Quarterly, 2 (3):299-318

Ni F., C. Su, S. Chung et K. Cheng (2005) “The Role of Trust in Supervisor in Budgeting System”, The 17th Asia-Pacific Conference on International Accounting Issues Wellington, New Zealand, Nov. 22

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 R.J. 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

Parker R.J. et L. Kyj (2006) “Vertical information sharing in the budgeting process”, Accounting, Organizations and Society, 31(1):27-45

Penno M. (1990) “Accounting Systems, Participation in Budgeting, and Performance Evaluation”, Accounting Review, 303-314

Roethlisberger F.J. et W.J. Dickson (1939) Management and the workers: an account of a research program conducted by Western Electric Company, Hawthorne Works, Chicago, Cambridge, Massachusetts: Harvard 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

Su J.T. et W.L. Lin “The effect of managerial performance in participative budgetary setting: information asymmetry as an intervening variable”, working paper.

Subramaniam N. et N.M. Ashkanasy (2001), “The Effect of Organisational Culture Perceptions on the Relationship between Budgetary Participation and Managerial Job-Related Outcomes”, Australian Journal of Management, 26(1)

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