



行政評価が予算編成過程に与える効果検証 : 常総市 ランダム化比較実験から

生方, 裕一
岡田, 幸彦
松尾, 貴巳

(Citation)

国民経済雑誌, 222(1):69-86

(Issue Date)

2020-07-10

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/E0042094>

(URL)

<https://hdl.handle.net/20.500.14094/E0042094>



国民経済雑誌

行政評価が予算編成過程に与える効果検証

—常総市ランダム化比較実験から—

生 方 裕 一
岡 田 幸 彦
松 尾 貴 巳

国民経済雑誌 第222巻 第1号 抜刷

2020年7月

神戸大学経済経営学会

行政評価が予算編成過程に与える効果検証

——常総市ランダム化比較実験から——

生 方 裕 一^a
岡 田 幸 彦^b
松 尾 貴 巳^c

本稿は、わが国の多くの地方自治体で事務事業を対象に実施されている行政評価に注目し、行政評価が次年度予算要求額に及ぼす効果を検証したものである。検証にあたり、茨城県常総市の一部の事務事業を対象に、次年度予算要求額を用いた予算要求額変化率、および次年度予算査定額を用いた予算査定額変化率を目的変数とするランダム化比較実験を行った。実験の結果から、行政評価が行われることで多様な予算要求が行われるという形で予算要求行動の変化をもたらす可能性が示唆された。加えて、減額査定が行われることを前提としつつも、質の高い予算要求が行われる上で、行政評価と実施計画の見直しが有用であることの可能性も示唆された。

キーワード 事務事業，行政評価，予算要求額，ランダム化比較実験

1 はじめに

行政評価は、わが国の地方自治体において「政策、施策及び事務事業について、事前、事中、事後を問わず、一定の基準、指標をもって、妥当性、達成度や成果を判定するもの」（総務省，2004）として自発的に取り組まれている「計画体系（政策—施策—事務事業）を中心とした業績管理システム」（松尾，2009，p.1）である。わが国では、New Public Management の考え方の普及に伴い、三重県での事務事業評価システムを筆頭に全国の地方自治体に広まっていった（梅田，2001；松尾，2009；山谷，2012）。2016年度現在、行政評価は、わが国の全都道府県，政令指定都市の95.0%，中核市の93.6%，特例市の97.3%，市区の83.5%に導入されている（総務省，2017）。

この導入の背景には、行政評価に対する財政効果への期待がある（古川，1999；湯下・坂

a 早稲田大学創造理工学部，y1_ubukata@aoni.waseda.jp

b 筑波大学大学院システム情報工学研究科，okayu@sk.tsukuba.ac.jp

c 神戸大学大学院経営学研究科，mats@kobe-u.ac.jp

野, 2002)。ところが実際には, 行政評価による予算編成時の内容的変化には至っていない(湯下・坂野, 2002)。総務省(2017)によると, わが国の地方自治体の中で, 行政評価を既に導入していると回答した1,099団体のうち71.3%が, 行政評価の課題として「予算編成等への活用」を挙げている。また, わが国の地方自治体では, 行政評価の結果が自明である事務事業の廃止・削減に用いられるのみで, 事務事業の改善や予算要求に用いられていない(田中, 2009)ことが指摘されている。

行政評価を予算編成に活用することの重要性と課題は既に先行研究¹⁾によって指摘されてきた(Lapsley, 1999; Cavelluzzo and Ittner, 2004; Pollitt and Bouckaert, 2017)が, 行政評価の予算編成に対する効果検証は長峯(2004)や松尾(2006)のような一部の研究を除き, あまり行われてこなかった。長峯(2004)は, 三重県における政策評価の効果検証を行っているが, 実務的に予算編成行動の中心となる事務事業レベルの行政評価は考慮されていない。一方で, 松尾(2006)は, 事務事業レベルの行政評価の貢献に注目してはいるものの, その焦点は当該事務事業の経済性・効率性・有効性にあり, 行政評価が現場の予算要求行動に対してもたらす効果を知る上では限界がある。総務省(2014)も, 行政評価の成果を問うてはいるが, 「個別の事務事業の効率性が向上した」や「予算配分を大きく変更できた」などのように主観的な判断に留まっている。

そこで本稿では, 事務事業レベルの行政評価が次年度予算要求・編成行動に対してもたらす効果を検証する。具体的には, 自治体の規模および行政評価の実情においてわが国の典型例と考えられる茨城県常総市において, 同市が担当する事務事業を対象としたアクション・リサーチ²⁾を, 中川(2006)が推奨するランダム化比較実験³⁾として行うことで検証した。これは, 行政評価自体が事務事業として行われており, 自治体の事務事業の効果検証には実験的手法が推奨される(中川, 2006)からである。

2 研究課題

2.1 行政評価の課題

行政評価は, 経営管理サイクル(Plan-Do-Check-Action; 計画—実行—評価—修正行動サイクル), いわゆるPDCAのマネジメント・サイクルを重視した経営管理システム(松尾, 2006; 松尾, 2009; 松尾, 2010)として実施される。わが国では, 各自治体の総合計画⁴⁾に従って実行されるものとしてこの経営管理システムが位置づけられてきた。

ここで, 本研究が注目する事務事業の行政評価とPDCAサイクル(松尾, 2006)を, 岡田ほか(2015)に基づいて整理する。岡田ほか(2015)では, 佐藤(2014)と松尾(2006)に基づき, Plan機能として実施計画・予算編成, Do機能として事務事業の実施・予算執行, Check機能として行政評価, Action機能として業務改善・実施計画改定(佐藤, 2014)を

それぞれ位置付けている。

このように、行政評価は「行政活動の成果を、自律的に計画（Plan）から実行（Do）、評価（Check）から改善（Action）へと連鎖させる一連のサイクルの構築のための手段」（坂元，2012，p.80）である。このPDCAサイクルが回ることによって各自治体の基本構想の実現が確実になる姿が、公共経営の1つの理想形であると想定される（岡田ほか，2015）。しかし、わが国の多くの地方自治体では、Checkを踏まえたAction、およびPlanとの連携が課題である。その原因として、宮本（2003）は、事務事業評価の単位と予算編成単位の一致が難しい⁵⁾ため、行政評価の結果が予算編成へと結びつきづらいつことを指摘している。また、コスト情報の不正確さ（田中，2009）や予算編成における政治的要因（Tassie et al, 1998；宮本，2003）も指摘されている。さらには、予算編成の議論が始まる夏時点までに、前年度分の事後的な評価情報の整理が間に合わず、結果に基づいた計画策定や予算編成が難しいこと（山本，2001）や、行政評価結果と予算配分を連動させると、組織は予算を確保しようとするために目標値を低く設定する可能性があること（上山・井関，2003）、コスト意識の向上には、現金主義ベースではなく発生主義ベースのコスト情報が必要であること（島田・三菱総合研究所，1999；自治省，2000；田中，2009）が指摘されている。

2.2 リサーチ・クエスチョン

このように行政評価の予算編成への活用には課題があるが、逆に、これら先行研究で指摘されてきた問題が解決されるのであれば、予算編成への活用、及び期待される効果が発揮される可能性が高くなる可能性がある。すなわち、行政評価と予算編成の事業単位が一致しており、当該事業に関する必要なコスト情報が提供されていること、さらには政治的な要因の影響を直接的に受け難い自治体内の財政部門による予算査定範囲内であれば、PDCAサイクルならではの効果が表れる可能性が高くなるといえよう。そこで本研究では、行政評価と予算編成の事業単位が一致していること、当該事業に関する必要なコスト情報が提供されていること、という2点を前提として、行政評価がPDCAサイクルにおいてもたらす効果を検証する。

PDCAサイクルのCheck機能に相当する行政評価に期待される効果の一つとして、事務事業担当者に対するCheck機能がある。これは行政評価を踏まえることで、事務事業担当者の予算要求行動が変化することを意味する。予算要求行動の変化が伴わなければ、影響システム（廣本，1986）としての行政評価に期待される業績管理システムとしての機能を期待することは難しいといえよう。そこで本稿では、予算要求額を予算要求行動の結果と見なし、行政評価の実施が予算要求額に与える影響に着目する。さらに、行政評価がCheck機能を果たし、当該PDCAサイクルが効果を発揮すると仮定すると、その効果は予算編成の初期

段階における財政部門における政治的価値判断を伴わない事務事業査定（湯下・坂野，2002）の結果に表れよう。なぜなら，Check 機能は過去の反省を促し，より効果的・効率的な事務事業の計画に結びつくと考えられるからである。したがって，予算要求額に対する予算査定額という点から，予算要求額の質にもたらす影響にも注目する。以上の点から，本稿における研究課題として，以下の2つのリサーチ・クエスチョン（RQ）を提示する。

RQ1 行政評価の実施は，予算要求額にどのような影響を与えるのか。

RQ2 行政評価の実施は，予算要求額の質にどのような影響を与えるのか。

3 ランダム化比較実験

3.1 常総市での旧行政評価

本稿では，上述の研究課題の検証にあたり，茨城県常総市で行われた2014年度事務事業評価対象とするランダム化比較実験を行った。常総市は，人口62,957人（2014年10月1日現在），面積123.52平方 km（2014年1月1日現在），財政力指数0.74（2013年度）の地方自治体であった。そして，田中（2009）が指摘する，①全庁的で画一的な手法，②事務事業評価，③業績測定，④業績測定以外の手法，⑤行政職員自身による実施，⑥評価シートの利用，というわが国自治体における行政評価の特徴に典型的にあてはまる行政評価を行ってきた。したがって，常総市は人口規模，及び財政面等から，わが国に存在する全1,788の地方自治体の中で代表的な中規模自治体⁷⁾と想定できること，および他の地方自治体と同様の行政評価を行っていたことから，自治体の規模および，基本的な行政評価の導入可能性という点で，行政評価の効果検証を行う上で適切な実験対象であると考えられた。

ランダム化比較実験においては，有効性があると考えている分析対象（本研究の場合は行政評価という業績管理システム）を適用群と非適用群に分けて適用しその有効性を検証する。分析対象にそもそも有効性がなければ差は生じない。本研究は，行政評価が次年度予算要求・編成行動に対してもたらす効果を検証することを目的としているため，常総市に適用される行政評価が有効性を持つ可能性があるかどうかについて予め検討しておく必要があった。

常総市では，2008年から行政評価が導入されていたが，予算編成方針では，漫然とした予算要求を止めることが課題として指摘され（岡田ほか，2015），それまでの行政評価は，予算要求・編成過程においてその有効性は認められていなかったため，本研究の検証にあたっては，有効性を高めるためのリフォームを行う必要があると認められた。そこで，次のような旧来の行政評価は一旦休止し，有効性が期待されるリフォーム後の行政評価について，ランダム化比較実験を行うこととした。

2013年以前の常総市の行政評価（以下，旧行政評価と称する）は，2008年から当時の企画

課によって本格導入⁸⁾された。旧行政評価は、 n 年度に $n-1$ 年度に実施された事務事業の善し悪しを評価し、これを $n+1$ から $n+3$ 年度（将来3ヶ年）についての実施計画のローリングに活かすとともに、 $n+1$ 年度の予算編成に活用する姿が想定されてきた。この行政評価は、行政運営の効率化と職員の意識改革にあり、これに併せて行政活動の成果向上やアカウントビリティなど多様なねらいを果たすために導入された。旧行政評価の対象は事務事業のみの1段階評価であり、公営企業会計を除き、企画課が選定した一部の事務事業を対象としてきた。そして、担当課を超えて評価を行うために課長補佐等が集う行政改革推進委員会、他部局の部長による1次評価と最終評価を決定するために市長・部長等が集う行政改革推進本部会議の2つの会における議論と合意形成に基づいて旧行政評価は行われてきた。旧行政評価では外部有識者による行政改革懇談会を開催して外部評価が行われてきたが、事務事業の可否等の公開評価は実施してこなかった。外部評価は内部評価をもとにしており、内部評価の対象全てについて住民代表としての行政改革懇談会が意見を論じ、併せてパブリック・コメントも求めている。

3.2 本実験研究で行った介入

2014年度は、前述の研究課題を検証するために旧行政評価の実施を中止した。そして、PDCA サイクルとしての費用対効果を高めることを目指し、旧行政評価と比較して、以下の6点について介入を行った。なお、それ以外の諸作業等については、旧行政評価と基本的に同様である。

まず1点目として、宮本（2003）が指摘し常総市でも生じていた事業単位の不一致問題の回避のために、Check 機能としての行政評価と Action 機能としての実施計画のローリングを、Plan 機能としての予算編成上の事業単位に統一した。2点目として、市の裁量や努力によって改善や変化の余地がある事業を全予算事業から行政評価対象事業として企画課と財政課が抽出した。3点目として、行政評価シートを、より予算とのつながりを意識させる方向で改良した。具体的には、旧行政評価におけるシートでは「今後の方向性」とぼやけていた評価ポイントを、本行政評価実験では「事務事業の内容（当初予算ベース）」と「予算への反映（当初予算ベース）」という2点から評価した。4点目として、事務事業担当者に対して、より真剣に行政評価を踏まえた予算要求が行われるよう、事務事業を担当し、行政評価シートを作成する係長等の名前を記載させた。5点目として、後述するPDCA 群とPDA 群について、ローリングした実施計画の提出を義務づけた。これは、PDCA サイクルの中での Check 機能としての行政評価の効果を検証するためには、Action 機能としての実施計画のローリングの状況に依存しない実験デザインが必要であると考えられたからである。最後に6点目として、財政課と企画課の協働（柴・松尾，2012）を目指し、各課の予算要求に

対する予算ヒアリングを財政課と企画課が協働して行った。

さて、常総市の予算編成方針では、漫然とした予算要求を止めることが毎年通告されてきた。これは、予算査定によって減額されることを想定しながら「これまでと同様に」という発想で工夫なく予算要求される現実に対する警告の意が込められているという（岡田ほか、2015）。ここで、減額査定が行われることを前提として、ある事務事業について工夫された予算要求が行われたと仮定した場合、その事務事業の予算要求はあまり減額されず、予算査定額が要求額に近い値となろう。すなわち、行政評価が行われることで、事業内容や予算額の再考に繋がり、予算査定時において予算査定者を納得させる予算要求が行われるものと考えられる。そこで、RQ2の検証にあたり、財政課長との意見交換を参考に、予算要求額の質を予算査定額変化率（＝予算査定額÷予算要求額）で測定することにした。これは、予算要求額に近い額で査定される場合、財政課にとって説得力の高い、すなわち質の高い予算要求が行われると考え、予算要求額に対して予算査定額が低ければ低いほど説得力の低い、すなわち質の低い予算要求が行われると考えられたためである。

なお、上述の2つのRQの検証にあたり、予算要求額への効果検証に影響を与えうるその他の要因を排除するため、一連の実験は厳格なブラインド実験として行われた。そのため、本実験研究の実施と検証は、実験デザインの構想を開始した2013年10月から各課が予算要求を終える2014年12月まで、各課には完全に秘密とした。さらに、RQ2については、財政課と企画課が関与する予算査定に関連することから、筆者らのみしか目的と方法を知らないという状況下で実験を行った。

3.3 対象事業の抽出とランダム・アサインメント

実験対象事業の抽出は、以下の過程を経て行われた。まず2014年度常総市予算書上の約300の事務事業のうち、2014年度当初予算において事業規模が100万円以上の継続事業であり、かつ1つの係が管轄している事業として160事業が抽出された。この160事業から、市に予算の裁量がない法定事業、法定事業ではないが査定できる範囲が極小であるとみなされた事業、来年度から予算額が大幅に増減する事業、今年度の予算額が大幅に増減した事業が除外され、92事業が本行政評価実験の対象事業として選択された。なお、この一連の候補事業選定は、常総市企画課と財政課が行い、筆者ら大学研究者は関与していない。

次に、筆者らは、選択された事業の4群へのランダム・アサインメントを行った。その際、行政評価実験の結果に影響を与えそうな事業の性質について、企画課と財政課と議論を行った。事業の予算規模、ソフト事業（主に役務提供を行うサービス事業）やハード事業（主に施設建設等の基盤整備に関わる事業）に代表される事業内容⁹⁾、さらには財源の多様性など様々な事業の性質が提示されたが、その中で事業の予算規模が最も実験結果に影響を与え

る要因であると考えられた。そこで本研究では、事業の予算規模の相違を事前に統制するために、層別ランダム・アサインメントによって群分けを行うことにした。その結果、上述の92事業が、2014年度当初予算の予算規模を基準に上位46事業と下位46事業とに分けられた。そして、これら2つをそれぞれ無作為に4群に群分けした。

この群分けを行った2014年6月時点では92事業が実験候補であった。しかし、実験中に2事業が指定管理者制度の導入、2事業が統廃合をすることに決まった。これらは本行政評価実験とは無関係な要因で決まったため、これら4事業は本論文における効果検証からは除いた。そのため、本論文の実証分析に用いる事業数は合計88事業となった。

本稿の実験研究は、この88事業に基づく4群を用いたランダム化比較実験である。具体的には、行政評価対照群として、行政評価を行いかつローリングした実施計画の提出を義務づける事業群（以下、PDCA群）と、行政評価のみを行う事業群（以下、PDC群）を用意した。提出先は、本実験研究の担当であった企画課とした。そして、行政評価非対照群としてローリングした実施計画の提出を義務づける事業群（以下、PDA群）と、行政評価非対称であり、かつローリングした実施計画の提出も義務付けない事業群（以下、PD群）を用意した（表1）。これによって、主分析としての行政評価対象事業群と非対象事業群との比較、そして詳細分析としてのPDCA群とPDA群との比較およびPDC群とPD群との比較を可能としている。

表1 実験群と事業数

実験群・事業数		行政評価		総数
		対 象	非対象	
実施計画の提出	義 務	PDCA群 (22事業)	PDA群 (22事業)	44事業
	義務ではない	PDC群 (22事業)	PD群 (22事業)	44事業
総 数		44事業	44事業	88事業

なお、この4群について、2014年度当初予算を目的変数とする一元配置分散分析を行ったところ、統計的な有意差は見られなかった（F値=1.080, p=0.362）。また、Kruskal-Wallis

表2 4群（2014年度当初予算）の基本統計量（千円）

事業群	平均値	標準偏差	最小値	第1四分位値	中央値	第3四分位値	最大値	標本数
全体	42,792	79,090	1,230	3,944	11,530	43,334	461,518	88
PDCA群	61,462	112,783	1,308	3,667	15,384	59,609	461,518	22
PDC群	21,065	21,710	1,230	3,961	13,213	34,744	71,838	22
PDA群	37,502	61,825	1,571	2,969	9,318	32,755	202,180	22
PD群	51,138	89,101	2,194	4,954	11,815	73,178	376,546	22

検定においても、同様に有意差は見られなかった ($p=0.763$)。このことから、88事業群に違いは見られず、無作為に4群に群分けされたと判断した(表2)。

4 実験結果

4.1 RQ1の検証結果

RQ1の検証にあたり、主分析として、予算要求額変化率(=2014年度次年度予算要求額/2014年度当初予算額)を目的変数として、行政評価対象群(PDCA群・PDC群)と行政評価非対象群(PDA群・PD群)による群間比較を行った。表3は、両群の基本統計量を示したものである。

表3 行政評価対象群・非対象群の予算要求額変化率

事業群	平均値	標準偏差	最小値	第1四分位値	中央値	第3四分位値	最大値	標本数
全体	1.762	2.061	0.324	1.017	1.197	1.566	12.593	88
行政評価対象群	2.124	2.724	0.325	1.019	1.243	1.916	12.593	44
行政評価非対象群	1.400	0.954	0.337	1.011	1.145	1.385	6.462	44

分散はF検定の結果から等分散とは考えられず(F値=7.322, $p=0.008$)、行政評価対象事業群にて、多様な予算要求が行われている状況が確認された。しかし、平均値については、t検定(t値=1.664, $p=0.102$)の結果から統計的な有意差は認められなかった。中央値についても、Mann-WhitneyのU検定($p=0.253$)の結果から統計的な有意差は認められなかった。そのため、多様な予算要求という形で工夫された予算要求が行われている傾向が観察されたが、その効果はあまり強いものではない可能性がある。

次に、詳細分析としてPDCA群とPDA群、およびPDC群とPD群とで群間比較を行った(表4)。

表4 4群の予算要求額変化率(n=22)

事業群	平均値	標準偏差	最小値	第1四分位値	中央値	第3四分位値	最大値
PDCA群	1.727	2.476	0.325	1.054	1.161	1.363	12.593
PDC群	2.521	2.955	0.831	1.002	1.441	2.467	11.970
PDA群	1.374	0.661	0.337	1.019	1.186	1.503	3.120
PD群	1.425	1.193	0.575	0.996	1.089	1.383	6.462

まずは、PDCA群とPDA群である。F検定の結果から、両群は等分散であると考えられた(F値=1.650, $p=0.206$)。また、平均値については、t検定(t値=0.644, $p=0.523$)の結果から、中央値については、Mann-WhitneyのU検定($p=0.842$)からそれぞれ統計的有

意差は認められなかった。

続いて、PDC 群と PD 群である。F 検定の結果から、両群は等分散とは考えられなかった (F 値=5.088, $p=0.029$)。そして、平均値については、t 検定 (t 値=1.613, $p=0.118$) の結果から、中央値については、Mann-Whitney の U 検定 ($p=0.105$) からそれぞれ統計的有意差は認められなかった。以上の結果から、予算要求額変化率に関して、詳細分析においても、主分析と同様の傾向が観察された。

ここで、PDCA 群よりも PDC 群の方が予算を多めに要求する傾向にあることに注目したい。この原因は、本行政評価実験の評価結果が PDCA 群と PDC 群とで偏った結果になったことに依存していると考えられる。PDCA 群の行政評価結果は、「事務事業の内容 (当初予算ベース)」も「予算への反映 (当初予算ベース)」も、全ての事業で「現行どおり」の結果であった。一方で、PDC 群は、「事務事業の内容 (当初予算ベース)」については「拡充」が22事業中10事業 (45%)、「予算への反映 (当初予算ベース)」については「増額」が22事業中2事業 (9%)、それ以外は全て「現行どおり」という、PDCA 群と比較して事業担当者の背中を後押しする行政評価結果となった。本行政評価実験はブラインド実験が徹底され、全行政評価の評価担当者はランダム化比較実験の群分けのことを全く知らない。つまり、サブグループ間での偶然の偏りが、PDCA 群と PDC 群との間の予算要求額の相違を生んだ可能性が高いものと考えられる。このことは、行政評価が事業担当者の予算要求行動に対して影響を与え (岡田ほか, 2015)、その結果、多様な予算要求という形で工夫された予算要求が行われる可能性があることを示唆している。

続いて、他の統制要因を考慮して上述の行政評価の効果を検証するために、予算要求額変化率 (=2014年度次年度予算要求額/2014年度当初予算額) を目的変数として、(1)式を推定した。

$$\begin{aligned} & \text{予算要求額変化率}_i \\ & = \gamma_0 + \gamma_1 \text{行政評価実施ダミー}_i + \gamma_2 \text{実施計画提出ダミー}_i \\ & \quad + \gamma_3 \text{2014年度当初予算}_i + \gamma_4 \text{ハード事業ダミー}_i + \gamma_5 \text{事業方向性ダミー}_i + \varepsilon_{it} \quad (1) \end{aligned}$$

(1)式は、説明変数として、行政評価実施ダミーと実施計画提出ダミーを用いている。行政評価実施ダミーは、行政評価対象事業であれば1、非対象であれば0とするダミー変数である。そして実施計画提出ダミーは、実施計画の提出が義務付けられていた事業であれば1、義務付けられていない事業であれば0とするダミー変数である。また、統制変数として、2014年度当初予算、ハード事業ダミー、事業方向性ダミーを用いた。2014年度当初予算は、実験対象となった88事業の割当の基準として用いたものである。さらに、施設管理等が主であるハード事業の場合、公的サービスの提供が主であるサービス事業と比べて予算要求が大きくなることが想定される。そこで、事務事業の特性を統制するために、施設管理等が主な事業

であるハード事業を1、それ以外を0とするハード事業ダミーを設定した。そして、行政評価の結果を統制するために、事業方向性ダミーを設定した。これは、行政評価対象事業44事業のうち、事業方向性¹⁰⁾について、拡充が10事業（PDC群10事業）、現行通りが34事業（PDCA22事業、PDC12事業）であったことを踏まえ、拡充の評価を受けた場合に1、現行通りの評価を受けた場合は0とするダミー変数である。

表5 分析に使用した変数の基本統計量 (n=88)

変数名	平均値	標準偏差	最小値	第1四分位値	中央値	第3四分位値	最大値
予算要求額変化率	1.762	2.061	0.325	1.017	1.197	1.566	12.593
予算査定額変化率 ¹¹⁾	0.796	0.255	0.009	0.686	0.866	0.959	1.424
行政評価実施ダミー	0.500	0.503	0.000	0.000	0.500	1.000	1.000
実施計画提出ダミー	0.500	0.503	0.000	0.000	0.500	1.000	1.000
2014年度当初予算	42791.590	79090.187	1230.000	3944.000	11530.000	43334.000	461518.000
ハード事業ダミー	0.300	0.459	0.000	0.000	0.000	1.000	1.000
事業方向性ダミー	0.110	0.319	0.000	0.000	0.000	0.000	1.000

推定にあたり、各変数の相関係数を確認した（表6）ところ、多重共線性が懸念される変数は確認されなかった。

表6 (1)式で用いる変数の相関係数 (n=88)

変数名	(1)	(2)	(3)	(4)	(5)	(6)
(1) 予算要求額変化率		0.123	-0.069	0.016	0.077	0.001
(2) 行政評価実施ダミー	0.177		0	-0.009	0	0.358
(3) 実施計画提出ダミー	-0.103	0		-0.038	0	-0.358
(4) 2014年度当初予算	-0.031	-0.019	0.085		0.201	-0.082
(5) ハード事業ダミー	0.204	0	0	0.171		-0.075
(6) 事業方向性ダミー	-0.055	0.358	-0.358	-0.121	-0.075	

(注) 左斜下は Pearson 相関係数, 右斜上は Spearman 順位相関係数を示している。

(1)式の OLS による推定の結果（表7）からは、行政評価実施ダミーがプラス有意であり、統制要因では、ハード事業ダミーがプラス有意であることが推定されている。このことは、行政評価対象事業、およびハード事業において予算要求額変化率の増加が見られることを示唆している。ハード事業は、施設管理等が主な事業であるため一般的に予算要求額が増加する傾向にあるが、表7の結果はその傾向を示している。また、統制変数を考慮しても、行政評価が予算要求額増加率に影響を与えることが示された。しかし、プラス有意であるため、行政評価による予算要求額の抑制効果については認められなかった。この結果は、行政評価実施群と非実施群とで事業費予算の伸びを比較し、非実施群より実施群の伸びが大きくなったとする松尾（2006）の結論に一致するものである。また、行政評価に期待される効果

は、予算抑制効果ではなく、次年度以降の事務事業の再考（山谷，2012）にある。そのため、本分析結果は、行政評価の実施が、前年度を踏襲した予算要求（黒川，2013）ではなく、将来を見据えて事業内容を再考した結果としての予算要求が行われる可能性があることを示唆している。

表7 重回帰分析の結果

変数名	モデル1	モデル2
定数	1.611 *** (4.273)	1.556 *** (3.740)
行政評価実施ダミー	0.724 * (1.663)	1.006 ** (2.181)
実施計画提出ダミー	-0.422 (-0.970)	-0.686 (-1.484)
2014年度当初予算		0.000 (-0.661)
ハード事業ダミー		0.906 * (1.908)
事業方向性ダミー		-1.267 (-1.617)
N	88	88
adj R ²	0.019	0.061

(注) (1)式についてOLSで推定した結果を示している。上の値が偏回帰係数，下の値が各説明変数のt値を示している。*，**，***はそれぞれ，10%，5%，1%水準以下で有意であることを示している。

4.2 RQ2の検証結果

RQ2の検証にあたり、予算査定額変化率（＝2014年度次年度予算査定額／2014年度次年度予算要求額）を目的変数として、行政評価対象群（PDCA群・PDC群）と行政評価非対象群（PDA群・PD群）による群間比較を行った。表8は、両群の基本統計量を示したものである。

表8 行政評価対象群・非対象群の予算査定額変化率

事業群	平均値	標準偏差	最小値	第1四分位値	中央値	第3四分位値	最大値	標本数
全体	0.796	0.255	0.009	0.686	0.866	0.959	1.424	88
行政評価対象群	0.774	0.282	0.043	0.595	0.862	0.972	1.424	44
行政評価非対象群	0.819	0.226	0.009	0.776	0.866	0.956	1.117	44

RQ2の検証においても、RQ1と同様に主分析と詳細分析を行った。分散はF検定の結果から等分散とは考えられなかった（F値＝4.436，p＝0.038），そして、平均値については、t検定（t値＝－0.833，p＝0.407）から、中央値については、Mann-WhitneyのU検定（p＝

0.520) から、それぞれ統計的有意差は認められなかった。

次に、詳細分析として PDCA 群と PDA 群、および PDC 群と PD 群とで群間比較を行った。まずは、PDCA 群と PDA 群である。両群は等分散である (F 値=0.203, $p=0.655$) と考えられた。そして、平均値については、t 検定 (t 値=0.472, $p=0.640$)、中央値については、Mann-Whitney の U 検定 ($p=0.833$) より、それぞれ統計的有意差は認められなかった。

続いて、PDC 群と PD 群である。両群は等分散とは考えられず (F 値=8.773, $p=0.005$)、平均値については、t 検定 (t 値=-1.507, $p=0.141$)、中央値については、Mann-Whitney の U 検定 ($p=0.557$) より、それぞれ統計的有意差は認められなかった。加えて、4 群の予算査定額変化率について、一元配置分散分析の結果 (F 値=1.496, $p=0.222$)、Kruskal-Wallis 検定の結果 ($p=0.782$) から、4 群間での違いに統計的有意差は認められなかった (表9)。

表9 4 群の予算査定額変化率 (n=22)

事業群	平均値	標準偏差	最小値	第1四分位値	中央値	第3四分位値	最大値
PDCA 群	0.848	0.211	0.477	0.695	0.881	0.956	1.424
PDC 群	0.700	0.326	0.043	0.485	0.731	0.995	1.093
PDA 群	0.815	0.251	0.009	0.773	0.879	0.987	1.117
PD 群	0.823	0.203	0.028	0.775	0.863	0.949	0.993

上記の観察結果を整理すると、全体として2割減の減額査定がなされており、本行政評価実験の対象となった全体88事業と行政評価対象事業群44事業との間には、予算査定額変化率、すなわち予算要求額の質の相違は見られなかった。ここで、行政評価対象群よりも非対照群の方が予算要求額の質が高い傾向にあることに注目したい。この点について、4群それぞれの予算要求額の質の違いを確認すると、平均値も中央値も統計的有意差は見られなかったとはいえ、行政評価対照群であった PDC 群よりも、行政評価非対象群であった PDA 群、PD 群の方が予算査定額変化率の平均値、および中央値が高い傾向にあることがわかる。このことは、予算要求額の質が、実施計画の見直しが義務付けられなかった PDC 群よりも高い傾向にあることを意味する。その理由の1つとして、行政評価の結果を踏まえた多めの予算要求がなされた PDC 群が、予算査定において大幅な減額査定を受けたことが考えられる。その結果、予算要求額変化率では平均値・中央値ともに4群で最大を示した PDC 群が、予算要求額の質では、平均値・中央値ともに最小となったといえよう。

PDA 群は、予算要求額変化率において、4群中で最小の標準偏差を示している。そのため、行政評価対照群と比べてあまり多様ではないと考えられる予算要求が見られた。しかし、PDA 群の予算要求額の質を踏まえると、質の高い予算要求が行われる上で、実施計画の見直しが有用である可能性がある。一方で、予算要求額の質に関して、PDA 群と同様の値を

示している PD 群であるが、同群の予算査定額変化率の最大値が 1 を下回っており、PD 群に属する全ての事業において減額査定されていることがわかる。

以上のことから、行政評価を実施することで工夫された予算要求はなされるものの (RQ1)、行政評価を実施するだけでは予算査定額変化率、すなわち予算要求の質は変わらない (RQ2)。しかし、実施計画の見直しも行うことで予算要求の質が高まる可能性がある (RQ2) と考えられる。この可能性は、行政評価と実施計画の見直しの両方を行った PDCA 群が、予算要求額の質に関して、平均値・中央値もいずれも 4 群の中で最高値であることから示唆される。

続いて、RQ1 と同様に他の統制要因を考慮したうえの効果検証を行うため、予算査定額変化率 (=2014年度次年度予算査定額/2014年度次年度予算要求額) を目的変数とする重回帰分析を行った。

予算査定額変化率_i

$$= \gamma_0 + \gamma_1 \text{行政評価実施ダミー}_i + \gamma_2 \text{実施計画提出ダミー}_i + \gamma_3 \text{2014年度当初予算}_i \\ + \gamma_4 \text{ハード事業ダミー}_i + \gamma_5 \text{事業方向性ダミー}_i + \gamma_6 \text{予算要求額変化率}_i + \varepsilon_{it} \quad (2)$$

モデル式(2)は、目的変数は予算査定額変化率に変更しているものの、RQ1 の検証で用いたモデル式(1)と同じ説明変数を用いている。ただし、予算査定額変化率を見ることから統制変数として、新たに予算要求額変化率を追加している。

推定にあたり、各変数の相関係数を確認した(表10)ところ、多重共線性が懸念される変数は確認されなかった。

表10 (1)式で用いる変数の相関係数 (n=88)

変数名	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) 予算査定額変化率		-0.069	0.083	0.019	-0.076	0.026	-0.535
(2) 行政評価実施ダミー	-0.089		0	-0.009	0	0.358	0.123
(3) 実施計画提出ダミー	0.137	0		-0.038	0	-0.358	-0.069
(4) 2014年度当初予算	-0.020	-0.019	0.085		0.201	-0.082	0.016
(5) ハード事業ダミー	-0.099	0	0	0.171		-0.075	0.077
(6) 事業方向性ダミー	0.013	0.358	-0.358	-0.121	-0.075		0.001
(7) 予算要求額変化率	-0.455	0.177	-0.103	-0.031	0.204	-0.055	

(注) 左斜下は Pearson 相関係数, 右斜上は Spearman 順位相関係数を示している。

(2)式の OLS による推定の結果(表11)からは、予算要求額変化率のみがマイナス有意であることが推定された。予算査定額は、予算要求額に対して減額されたものとなる¹²⁾ことが想定されるが、想定通り予算要求額変化率の係数が負となっている。また、行政評価実施ダミーに有意差が認められなかったことから、他の統制要因を考量したとしても、行政評価を実施することで工夫された予算要求はなされるものの (RQ1)、行政評価を実施するだけでは予算査定額変化率、すなわち予算要求の質は変わらない (RQ2) ものと考えられる。

表11 重回帰分析の結果

変数名	モデル1	モデル2
定数	0.784 *** (16.716)	0.874 *** (16.619)
行政評価実施ダミー	-0.045 (-0.839)	-0.012 (-0.215)
実施計画提出ダミー	0.070 (1.283)	0.054 (0.984)
2014年度当初予算		0.000 (-0.390)
ハード事業ダミー		0.000 (-0.005)
事業方向性ダミー		0.024 (0.261)
予算要求額変化率		-0.054 *** (-4.208)
N	88	88
adj R ²	0.004	0.160

(注) (1)式についてOLSで推定した結果を示している。上の値が偏回帰係数、下の値が各説明変数のt値を示している。*, **, ***はそれぞれ、10%、5%、1%水準以下で有意であることを示している。

ただし、この行政評価がもたらす効果は、評価結果によって変わりうるものである。また、予算要求時の活用の仕方によっても変わり得る（岡田ほか、2015）。加えて、行政評価を業績管理システムの一つと見なし、この業績管理システムについて、Simons（1995）が示した診断的コントロールとインタラクティブコントロールに着目した研究が増えている（Bisbe et al. 2007; Widener 2007）。そのため、行政評価情報の活用の効果をより詳細に検討する際には、制度としての行政評価に着目するのではなく、評価を受けてどのように活用したのか、すなわち行政評価情報の活用方法ごとの効果を分析する必要がある（目時・妹尾、2012）。

本稿の実験研究では、行政評価の予算編成時の活用について、行政評価、実施計画のローリング後に各事業担当者に対する個別ヒアリング調査を行っている。その内容については、紙面の都合上、別稿で扱うものとする。

5 おわりに

本稿は、わが国地方自治体における業績管理システムとしての行政評価について、その予算編成への影響を実験的に検証した。検証の結果、統計的有意差は認められなかったものの、行政評価が行われることで、多様な予算要求が行われる可能性が示唆された。さらに行政評

価と実施計画の見直しを組み合わせることが、予算査定との関係で予算要求額の質を高める可能性も示唆された。

本稿の実験研究は、常総市を対象とした1事例ではあるが、本結果は、典型的な中小規模自治体における事例として、行政評価の効果の再現性や行政評価システムの再構築可能性が高いものと考えられる。しかし、本研究結果は、2014年度の当初予算のみに基づいた割当の結果であり、かつ予算編成との連携を阻害する要因の影響を直接受け難いと考えられた一部の事務事業に基づくものである。そのため、対象事業の選択方法や群分けの基準について、本実験研究で用いた当初予算基以外にも存在する可能性がある。それらを用いた対象事業の抽出や群分け、さらには、各事務事業の内容や総合計画、首長の方針などの上位の計画等を考慮に入れた分析が今後の課題であるといえよう。

注

- 1) 主にアングロサクソン系諸国では、1980年代以降の New Public Management に基づき、行政評価結果を活用して内部管理を行う組織が志向された (Pollitt and Bouckaert, 2017)。そして、Cavelluzzo and Ittner (2004) は、幹部職の寄与、意思決定権限、業績管理手法の習熟の程度が、当該組織において、結果志向 (result-oriented) の業績管理の促進・阻害要因となりうることを示している。その一方で、Lapsley (1999) は、行政組織の活動成果を定量的に測定し、評価することの困難性も指摘されている。
- 2) アクション・リサーチは、専門的な研究者と研究対象である組織のメンバーが一つの共同的なチーム (もしくはネットワーク) を形成し、研究対象メンバーの置かれた状況を改善するための研究アプローチ (Greenwood and Levin, 2007) と定義される。
- 3) ランダム化比較実験は、「ランダム割り付けによって、2つまたはそれ以上の数の群を作る。1つまたはそれ以上の数の群に対してはある介入を行い (実験群)、他の群 (比較群またはコントロール群) に対しては別の措置を行うか、何もしない。介入の効果は、両群に生じた結果を比較することによって観察する。もし、ランダム化によってできた群のサイズが十分に大きなものであれば、両群の間に観察された差は、介入以外の、既知あるいは未知の変数の影響によるものではなく、介入の結果であることに確信が持てる」 (Torgerson and Torgerson, 2010, p.2) 実験手法である。
- 4) 総合計画とは、営利企業でいう中長期経営計画に相当し、旧地方自治法第2条第4項を根拠に策定される基本構想、基本計画、実施計画の3層構造からなる。
- 5) 単位の一一致が難しい理由として、地方財政法等の法律において、事務事業よりかなり粗い事業区分を予算編成単位とすることが容認されているため、予算編成単位として、事務事業より粗い単位や組織別など単位で設定することが可能になっているためである (松尾, 2006)
- 6) 常総市では予算書上の事業として、事務事業を対象に事務事業評価が行われている。同市の予算書上では事務事業評価という名称であり、その内容は一般的な行政評価と同様である。
- 7) 「人口階級別市町村数」(総務省統計局, 2015) によれば、常総市の人口に該当する階級 (5万人以上10万人未満) に含まれる市数の全国市数に対する割合は33.8%であり、10万人以上の階級

- に含まれる市数の割合(34.1%)、5万人未満の割合(32.1%)からすれば、常総市の人口規模は全国の市の中で中間的な層に位置する。
- 8) 2013年10月1日時点では、常総市と同様に、 $n-1$ 年度の事務事業の評価結果を $n+1$ 年度予算の編成過程で用いる市区町村は全体の84.4%であった(総務省, 2014)。
- 9) 「一般的に事業の種類は、『新規・継続』『ハード(公共事業)・ソフト(公共事業以外の事業)』の2軸によって、①継続のソフト事業、②新規のソフト事業、③継続のハード事業、④新規のハード事業の4つに大別できる」(佐藤, 2013a, p. 57)。
- 10) もう一つの評価項目である予算方向性については、行政評価対象44事業のうち、2事業が増額で42事業が現行通りであった。この分析では、行政評価非対称群(PDC群, PD群)の44事業については全て0に設定した。そのため、予算方向性は、88事業中2事業のみが増額という事になり、標本数に課題があるという点からコントロール変数として用いなかった。
- 11) 予算査定額変化率は次節にて扱う。本稿の執筆にあたり、紙面が限られていることと、同じ変数を用いることから、当該変数の基本統計量をここに記載する。
- 12) 常総市財政課と企画課によると、常総市における例年の状況として、予算要求額は全体として増額申請される傾向がある。この予算要求に対して、財政課による予算査定は当該事務事業の必要性や経済性・効率性・有効性などが審査され、減額を基本とする査定額の決定がなされてきた。

参 考 文 献

- Bisbe, J., Batista-Foguet, J. M., and Chenhall, R (2007). "Defining management accounting constructs: A methodological note on the risks of conceptual misspecification." *Accounting, organizations and society*, Vol. 32, No. 7&8, pp. 789-820.
- Cavalluzzo, K. S., and C. D. Ittner (2004). "Implementing performance measurement innovations: evidence from government," *Accounting Organizations and Society*, Vol. 29, No. 3, pp. 243-267.
- Greenwood, D. and M. Levin (2007). *Introduction to Action Research: Social Research for Social Change, 2nd ed*, Sage Pub. Inc, CA.
- Lapsley, I (1999). "Accounting And The New Public Management: Instruments Of Substantive Efficiency Or A Rationalising Modernity?" *Financial Accountability & Management*, Vol. 15, No. 3&4, pp. 201-207.
- Pollitt, C., and Bouckaert, G (2017). *Public management reform: a comparative analysis-into the age of austerity.* Oxford University Press.
- Simons, R. (1995). *Levers of control: How managers use innovative control systems to drive strategic renewal.* Boston, MA: Harvard Business School Press.
- Tassie, B., Murray, V., and Cutt, J (1998). "Evaluating social service agencies: Fuzzy pictures of organizational effectiveness," *International journal of voluntary and nonprofit organizations*, Vol. 9, No. 1, pp. 59-79.
- Torgerson, D. J., and Torgerson, C. J (2008). *Designing randomised trials in health, education and the social sciences: an introduction.* Springer. (原田隆之・大島巖・津富宏・上別府圭子訳. 2010. 『ランダム化比較実験(RCT)の設計 ヒューマンサービス, 社会科学領域における活用のために』日本評論社)

- Widener, S. K (2007). "An empirical analysis of the levers of control framework." *Accounting, organizations and society*, Vol. 32, No. 7&8, pp. 757-788.
- 上山信一・井関友伸 (2003). 『自治体再生戦略—行政評価と経営改革—』日本評論社
- 梅田次郎 (2001). 「業績測定型・三重県事務事業評価システムの発展過程と展望」『日本評価研究』第1巻第2号, pp. 69-77.
- 岡田幸彦・下條雅弘・松尾貴巳 (2015). 「行政評価の効果検証：常総市におけるランダム化比較実験」第74回日本会計研究学会報告論文.
- 黒川和美 (2013). 「官僚行動の公共選択分析」株式会社勁草書房.
- 坂元英毅 (2012). 「業績情報の予算編成への活用に関する考察—わが国地方自治体における行政評価の実態に基づく検討—」『経営戦略研究』第6号, pp. 73-90.
- 佐藤幹 (2014). 「自治体予算のマネジメント・コントロール・システム化をめぐる問題」『会計ブロードグレス』第15号, pp. 38-48.
- 佐藤徹 (2013a). 「行政経営システムにおける行政評価の構造と機能」『高崎経済大学論集』第56巻第2号, pp. 53-66.
- 佐藤徹 (2013b). 「行政経営システムにおける総合計画の構造と機能」『地域政策研究』第16巻第1号, pp. 15-32.
- 自治省 (2000). 「地方公共団体の総合的な財政分析に対する調査研究会報告書」.
- 柴健次・松尾貴巳 (2012). 「行財政改革のための「予算企画」」『会計検査研究』第46号, pp. 45-63.
- 島田晴雄・三菱総合研究所政策研究部 (1999). 『行政評価』東洋経済新報社.
- 総務省 (2004). 『平成16年度 地方公共団体における行政評価の取組状況』.
- 総務省 (2014). 『地方公共団体における行政評価の取組状況等に関する調査結果』.
- 総務省 (2017). 『地方公共団体における行政評価の取組状況等に関する調査結果』.
- 総務省統計局 (2015). 『第六十四回日本統計年鑑 平成27年』.
- 田中啓 (2009). 「日本の自治体の行政評価」『分野別自治制度及びその運用に関する説明資料』No. 14.
- 中川雅之 (2006). 「政策の実験的手法」『計画行政』第29巻第3号, pp. 3-10.
- 長峯純一 (2004). 「地方分権・政策評価による資源配分のガバナンス」『ファイナンシャルレビュー』5月号, pp. 59-78.
- 廣本敏郎 (1986). 「わが国製造企業の管理会計—1つの覚書—」『ビジネス・レビュー』第33巻第4号, pp. 64-77.
- 古川俊一 (1999). 「公共部門評価システムの制度化と限界—計画行政への含意—」『計画行政』第22巻第4号, pp. 8-13.
- 松尾貴巳 (2006). 「地方公共団体における業績評価システムの導入研究—予算編成における行政評価システムの意義とその効果についての実証分析—」『会計検査研究』第33号, pp. 121-135.
- 松尾貴巳 (2008). 「地方自治体における行政評価の有効性に関する実証研究」『公会計研究』第10巻第1号, pp. 37-56.
- 松尾貴巳 (2009). 『自治体の業績管理システム』中央経済社.
- 松尾貴巳 (2010). 「自治体における業績評価システムの多様性と有効性」『国民経済雑誌』第202巻第2号, pp. 29-45.

- 宮本幸平 (2003). 「自治体業績評価におけるフィードバックの諸問題」『会計』第164巻第9号, pp. 91-102.
- 目時壮浩・妹尾剛好 (2012). 「公会計・行政評価情報の行政経営への活用に向けた課題：混合研究法に基づく考察」『原価計算研究』第36巻第2号, pp. 115-129.
- 山本清 (2001). 『政府会計の改革』中央経済社.
- 山谷清志 (2012). 『政策評価』ミネルヴァ書房.
- 湯下健一・坂野達郎 (2002). 「行政評価システム導入県における目的志向型行政運営の実態と評価：三重県を事例として」『計画行政』第25巻第1号, pp. 109-117.