

SUMMARY EXECUTION: THE IMPACT OF ALTERNATIVE SUMMARIZATION STRATEGIES ON LOCAL GOVERNMENTS

JOSEPH DREW

BRIAN DOLLERY

University of New England

Joseph Drew is a Research Fellow in the Centre for Local Government at the University of New England. His research interests focus on performance measurement, corporate governance, expenditure and revenue structures for local government. He has consulted with numerous Victorian and New South Wales councils on municipal reform, accounting, finance and economic matters. Joseph has also been called as an expert witness for State and Federal government Upper House inquiries.

Brian Dollery is Professor of Economics and Director of the Centre for Local Government at the University of New England. Recent books include *Funding the Future* (2013), *Councils in Cooperation* (2012), *Local Government Reform* (2008), *The Theory and Practice of Local Government Reform* (2008), *Reform and Leadership in the Public Sector* (2007) and *Australian Local Government Economics* (2006).

ABSTRACT

Performance management in the public sector, including local government, has become far more pervasive in recent decades. Often performance indicators are summarized into a single score to enhance understanding and ease dissemination. However, the summation of performance indicators carries a risk that the rating assigned may largely be an artefact of the summarization strategy rather than an accurate representation of municipal performance. We employ the recent evaluation of New South Wales' municipal performance to demonstrate that the performance indicator compilation strategy is indeed a major determinant of the ratings assigned to local councils. Moreover, we illustrate how ratings may

exert a constitutive effect on municipalities by altering organizational behavior. A number of policy lessons are drawn from our empirical analysis, including significant methodological considerations and the need for higher levels of transparency.

Keywords: Constitutive accounting theory; local government; performance indicators

INTRODUCTION

Although performance management in the public sector, including local government, has a long history (Van Dooren et al. 2010; Moynihan 2008), over the past three decades it has become much more pervasive (see, for example, Pollitt and Bouckaert 2000). No doubt this is partly related to the rise of New Public Management (NPM) strategies over the same period. Indeed, it has been claimed that performance management systems are the ‘engine room’ of NPM (Diefenbach 2009). The ubiquitous embrace of performance measurement and assessment in the public sector has led to growing unease amongst public finance scholars. Moreover, the heavy emphasis on ‘management by numbers’ in public sector performance measurement and assessment in particular has attracted substantial critical attention (see Andrews et al. 2011 for a synoptic review of this literature).

Often performance indicators are summated into a single rating score to enhance understanding, ease dissemination (particularly by the media), and provide a handy aggregate measure of a local council’s performance (Saltelli 2007). However, the process of combining performance indicators into a single rating score has attracted substantial criticism. Principal objections include a loss of information, such as the loss of relative standing within broad categorical bands (Saisana et al. 2005), forfeiture of measures of uncertainty (Bird et al. 2005), the illusion that categorization is free of value judgements (Saltelli 2007; Bird et al. 2005; Kloot and Martin 2000), and the distortion of information when independent dimensions are combined (Bird et al. 2005). This last concern is particularly important in the context of municipal indicators since it is clear that performance occurs along a number of dimensions (Drew and Dollery 2015). For

instance, local government has a role in infrastructure management, particularly roads, but it also provides services to local residents, such as aged care and child care. Moreover, there are dimensions of municipal performance – such as the quality of customer service – which may defy quantification altogether (Kelly and Swindell 2002). Thus, when municipal performance data is reduced to a single number there is a substantial danger of one dimension being amplified at the expense of other dimensions of performance.

Despite the volume of scholarly attention directed at performance indicators, much work remains to be done, especially on composite performance indicators. For example, Jacobs and Goddard (2007, 108) have pointed out that ‘there is a paucity of research on how these composite performance indicators are constructed, what the methodological challenges are in doing so, and whether they are in fact a good reflection of performance’. The present paper seeks to address this gap in the literature by applying a number of frequently used summarization algorithms (or indexes) to a sample of New South Wales (NSW) local authorities to compare the outcomes of the different techniques. In so doing, we seek to demonstrate that the compilation algorithm employed is in fact a major determinant of the performance rating assigned to a given municipality. The constitutive implications for individual entities are also important and a broader message arising from this paper is that the performance management systems which lie at the heart of NPM are not objective ‘facts’ but rather the outcomes of a myriad of subjective decisions relating to how accounting data is compiled.

By way of institutional background, NSW local government is in the throes of a vigorous debate over structural reform through compulsory council consolidation engendered by the establishment of an Independent Local Government Review Panel (ILGRP) in 2012 by the NSW Government. The Panel recommended the amalgamation of more than 40 percent of the 152 NSW councils citing the need to improve financial sustainability. A key document informing the municipal merger recommendations was a NSW Treasury Corporation (TCorp) (2013) report entitled *Financial Sustainability of the New South Wales Local Government Sector*. In this Report

TCorp (2013) developed ten Financial Sustainability Ratios (FSR)¹ which it aggregated into a single Financial Sustainability Rating (see Table 1).

It is noteworthy that the performance assessment of NSW councils was based entirely on financial statement data. This is problematic since it means that other dimensions of performance, such as service quality, have been largely neglected (Drew and Dollery 2014). Moreover, financial data is necessarily orientated towards the past rather than future ability to provide municipal services. These matters raise questions regarding the process of evaluating the performance of NSW councils. However, the principal focus of the present paper revolves around a matter of much broader applicability: whether the method chosen to summarize performance indicators is a determinant of the rating or ranking that a council receives. If this is indeed the case, then it would in large measure undermine the claim that performance ratings, such as the ratings assigned by TCorp (2013), are an objective assessment of a council's performance.

The paper is divided into five main parts. Section 2 provides a synoptic account of constitutive accounting theory which is used to interpret the outcomes arising from differing composite performance measures. Section 3 describes six linear summarization strategies frequently employed in the empirical literature on local government performance assessment. Section 4 applies these six summarization algorithms (or indexes) in both unweighted and weighted contexts and compares the results with the TCorp (2013) FSR. Section 5 considers the constitutive implications of alternate summary methods for two exemplar councils. The

¹ TCorp (2013, 5) employed the following definition to inform the FSR: 'A local government will be financially sustainable over the long term when it is able to generate sufficient funds to provide the levels of service and infrastructure agreed with its community'. We acknowledge that the definition of sustainability is narrow and that this places limitations on our study. Moreover, the exclusive use of financial data in compiling FSR excludes other important dimensions of sustainability, such as citizen satisfaction and the property tax burden (see Drew & Dollery, 2014 for a discussion on the limitations of the approach taken by TCorp (2013)).

paper ends in Section 6 by discussing the public policy implications of aggregating performance indicators into composite ratings.

CONSTITUTIVE ACCOUNTING THEORY AND FINANCIAL SUSTAINABILITY RATINGS

Hopwood's (1987) constitutive accounting theory represents a radical departure from the traditional concept of accounting systems as 'value-free' reflections of an organization (Cunningham and Harris 2005). By employing three case studies on the evolution of organisations and their accounting systems, Hopwood (1987, 229) concluded that 'by creating quite particular objectifications of the otherwise vague and abstract, and particular conceptions of economic facts, accounting also can create not only a context in which the conditions exist for other organizational practices to change but also a means by which a particular organizational visibility can compete for or be imposed upon managerial attention'. Accordingly, far from providing a neutral portrayal of an organization, accounting may actually 'reshape' the organization over time. This paper examines how financial sustainability ratings (FSR) - in the present case a summation of ten financial ratios - act as a constitutive agent consistent within Hopwood's (1987) theory.

The FSR assigned by NSW TCorp (2013) are accounting 'facts created by the craft [that] gives rise to an influential language and set of categories for conceiving and changing the organization' (Hopwood 1987, 229). Indeed, the FSR classification categories of 'distressed', 'very weak', 'weak', 'moderate', 'sound', 'strong' and 'very strong' (TCorp 2013) are designed to convey a relative index of the financial sustainability of all NSW local authorities. They have been cited as 'economic facts' to justify the proposed compulsory consolidation of 63 municipalities into 20 new entities in the Australian state of NSW, as recommended by the ILGRP (2013b)². Moreover, their influence on policymaking can

² We acknowledge that it is surprising for a restricted set of financial ratios to be employed for municipal merger decisions. However, there is precedent in Australia for this practice. For instance, the Queensland Local Government Reform Commission

hardly be exaggerated. For example, the FSR appear in the *Comparative Information on NSW Local Government* report (NSW Division of Local Government 2013), dozens of media reports, and academic papers, such as Drew and Dollery (2014).

Yet summary ratings, like the FSR, may well have an impact beyond policy making and performance appraisal: they can also function as a pedagogical instrument by ‘define[ing] the frameworks people think and act within, what they are striving for, how they are being evaluated, and how they behave and even what they become’ (Diefenbach 2009, 900). Furthermore, the influence of the new pedagogic discourse will be aided by the apparent simplicity of the ‘objectifications of the otherwise vague and abstract’ (Hopwood 1987) embodied in FSR: terms such as ‘weak’ or ‘strong’ do not require the mastery of abstract accounting concepts, thereby facilitating FSR dissemination outside of the usual management circles.

It is not unreasonable to suggest that the aggregated effects of this pedagogical process will exert an existential influence on the organization itself through the imposition of ‘a particular organizational visibility’ (Hopwood 1987). For instance, a council classified as ‘very weak’ is likely to become an expenditure focused entity: new initiatives or service improvements which materially increase expenditure will diminish and previously inconceivable proposals, like increasing property taxes and fees and charges, could gather support. The fact that a negative financial sustainability assessment may spur a council to take determined action to improve its financial position is no bad thing. The question at the heart of this paper is whether the financial sustainability rating given to a council can be considered to be an ‘economic fact’ or whether it is instead an artefact of the summarization method employed.

There is general agreement in the literature that ‘composite measures may send misleading, non-robust policy messages if they are poorly constructed or misinterpreted’ (Saltelli

also employed a small set of financial ratios to inform its decision making which led to a reduction in Queensland municipalities from 157 to just 73 (see, for example, Drew & Dollery, 2013).

2007, 69). It is thus clear that plausible grounds exist for evaluating the objectivity of 'economic facts' which are created by accounting practices, such as the FSR. Indeed, all performance compilation algorithms (or indexes) rest on implicit value judgements, weights and trade-offs which represent not only a technical method to summarize Performance Indicators, but also possess a 'political' dimension (Diefenbach 2009). The possibilities for arriving at an overall measure of financial sustainability are limited only by the analyst's imagination and it would thus be incorrect to assume that the 'facts' created by the craft are objective and value free. In section 3 we evaluate six prominent methods drawn from the relevant literature in order to demonstrate a mere subset of the possibilities which are available to local government performance analysts. We then proceed to apply each method to 2011 NSW data – where the FSR were based on 2011 performance indicators published by TCorp in 2013 - before considering the different ways in which municipalities might be respond to equally valid alternative objectifications of accounting material otherwise 'vague and abstract'.

SUMMARIZATION STRATEGIES

In the *Handbook of Local Government Fiscal Health*, Maher and Deller (2013) cite quartile ranking (Brown 1993), binary ranking (Kloha et al. 2005), cluster analysis (Zafra-Gomez et al. 2009), principal components analysis (PCA) and factor analysis (FA) (Congressional Budget Office 1978) as methods which can summarize a range of financial ratios and other performance indicators into a single measure of fiscal health. In addition, in a review of general health indicators, Hendrick (2004) cites examples of standardization (also known as normalization or z-scoring, including Nathan and Adams (1976) and the U.S. Department of Treasury (1978)). Finally, in the European Commission Joint Research Centre's *Tools for Composite Indicators Building* Nardo et al. (2005) discusses Principal Components Analysis (PCA), Factor Analysis (FA), cluster analysis, standardization, and re-scaling as linear methods for the construction of summary categorizations of performance indicators.

However, these methods represent a mere subset of the infinite possibilities confronting performance analysts. In

fact, non-linear techniques, such as Data Envelopment Analysis (DEA) and geometric aggregation, are also possible. However, in this paper we obviously cannot survey every possible method. Nor is this necessary to demonstrate that the way in which performance indicators are summarized into a single composite measure can have significant constitutive implications for individual municipalities. We thus focus on the principal methods used in the extant literature: standardization, scaled summation, binary scoring, quartile scoring, FA and PCA. We have elected to avoid an application of cluster analysis owing to the fact that there are a myriad of alternate approaches, including linkage, centroid, Ward's, k-means, k-median, which are subject to a number of permutations based on similarity measure (such Euclidean distance or angular separation) and cut-off rules. Moreover, there is no clear way of ranking the final clusters that emerge.

Standardization is a method by which different distributions of scores are transformed into a common scale with a mean of zero and a standard deviation of one. It is a simple process for which most contemporary statistical packages have well established routines. Furthermore, the standard normal distribution is often used as the basis for hypothesis testing and it is thus well-known to most analysts. However, the standard transformation rests on a crucial assumption that the performance indicators are normally distributed, which is unlikely to be the case for each and every performance indicator comprising a given summary rating. By way of contrast, range scaled summation transforms performance indicators by scaling them according to the range of observations. This method requires no *a priori* knowledge of the performance indicator distribution and is relatively resistant to the distortionary effects of outliers. Both methods are highly accessible and suitable for weighting of individual performance indicators.

Binary scoring (Kloha et al. 2005) assigns an integer for a given performance indicator which achieves the benchmark condition and a lower value where the performance indicator does not achieve the benchmark. Its validity thus rests on the

soundness of designated benchmarks³. Kloha et al. (2005, 317) suggest that some benchmarks are straightforward and logical, like the net fund balance, whilst others should be constructed by identifying a small percentage which are ‘standard deviations from average values’. In this instance, we have elected to use the TCorp (2013) benchmarks given that (a) these appear to have been used in the formulation of TCorp FSR (although the process of how the benchmarks and weightings of performance indicators have been summarized is not publicly documented) and (b) the Kloha et al. (2005) process lacks detail and may incorrectly assume performance indicators follow a normal distribution. Binary scoring is a simple process once benchmarks have been assigned. However, it (a) side steps theoretical and technical problems associated with assigning suitable benchmarks and (b) lacks definition owing to the binary assessments of each performance indicator (generally only providing a range of rankings for a given municipality). On the other hand, quartile scoring (see, for instance, Brown 1989; Zafra-Gomez et al. 2009) provides more definition in addition to using measures of central tendency which are not skewed by outliers, such as means and standard deviations. However, this approach would seem to condemn a quarter of the councils simply by virtue of their *relative* position, irrespective of their *absolute* performance (Kloha et al. 2005).

By way of contrast, FA is a sophisticated statistical technique which utilizes the *covariance* between individual observations (councils) to hypothesize latent linear composite causal factors (in this case a hypothetical sustainability factor). The principal factors thus adduced are then used to (in this instance) produce a single number for each municipality, which represents a reduction of the performance indicators (see Kim & Mueller (1978) for further details). Principal components analysis differs in its attempt to explain the maximum *variance* possible within the data by summarizing it as linear combinations of the observations. Unlike FA, PCA does not depend on a

³ Specifically, we assigned the value of 1 when the benchmark was achieved and a value of 0 when it was not. We then sum the scores and rank councils on the basis of the summation (councils are ranked in descending order).

hypothetical causal model (see, for instance, Duntelman (1989)). Both techniques require no distributional assumptions, but nonetheless can be sensitive to the presence of outliers. One common approach to this problem is to exclude extreme outliers. However, the resultant truncated data set is then of little value to regulators (Nardo et al. 2005). A number of robust estimators have been proposed in the literature. However, 'of the [many] robust procedures available, no single method works best in all situations' (Zygmunt and Smith 2014). Because of this uncertainty we have elected not to present alternative FA and PCA estimations which attempt to deal with the presence of outliers. Suffice to say that the potential of outliers to distort these summation techniques, and the significant variation arising from diverse methods to deal with any distortion, simply adds another critical decision to the performance indicator summation conundrum.

Finally, TCorp FSR are weighted summaries of the ten performance indicators defined in Table 1. Unfortunately TCorp (2013) has not disclosed exactly how these performance indicators are combined, although benchmarks and weightings are detailed in *Financial Sustainability of New South Wales Local Government Sector*. The lack of agreement between FSR and the six linear summary methods detailed in Table 2 and Table 4 suggest that TCorp in all probability has not drawn on any of the principal methods detailed in the scholarly literature. It is thus critical that TCorp explain the summarization algorithm so that scholars and municipal managers alike can have confidence in the FSR assessments. However, for the purposes of this paper, it is only necessary to demonstrate that different summary methods (including different methods for dealing with outliers) can produce very different ratings, which may result in significant constitutive consequences for local government entities over time.

Table 1
Definitions, Benchmarks and Weightings of TCorp
Financial Sustainability Ratios

Variable	Weighting	Benchmark	Definition	Median
Dependent				
Operating ratio	17.5%	>-4%	(operating revenue ^a - operating expenses) / operating revenue ^a .	-6.15
Own Source Revenue ratio	17.5%	>60%	rates, utilities and charges / total operating revenue ^b .	58.85
Unrestricted Current ratio	10.0%	>1.50x	current assets less restrictions / current liabilities less specific purpose liabilities.	3.05
Interest Cover ratio	2.5%	>4.00x	EBITDA ⁴ / interest expense.	13.8
Infrastructure Backlog ratio ⁵	10.0%	<0.02x	estimated cost to bring assets to a satisfactory condition / total infrastructure assets.	0.085
Debt Service Cover ratio	7.5%	>2.00x	EBITDA / (principal repayments + borrowing costs).	5.74
Capital Expenditure ratio	10.0%	>1.10x	annual capital expenditure / annual depreciation.	1.02
Cash Expense ratio	10.0%	>3.0 months	(current cash and equivalents / (total expenses - depreciation - interest costs)) x 12.	3.4
Buildings and Infrastructure Renewal ratio ⁶	7.5%	>1.00x	Asset renewals / depreciation of building and infrastructure assets.	0.59

⁴ EBITDA (earnings before interest expense, depreciation, tax and amortisation) is an acronym commonly employed in accrual based accounting systems, such as in Australia.

⁵ The Infrastructure Backlog ratio is a measure of the cost to bring assets up to a satisfactory standard expressed as a proportion of the value of the asset base. This data is based on engineering estimates and it is included in the special schedules appended to the financial statements.

Asset Maintenance ratio	7.5%	>1.00x	actual asset maintenance / required asset maintenance.	0.86
-------------------------------	------	--------	--	------

^a revenue excludes capital grants and contributions

^b revenue includes capital grants and contributions

COMPARISON OF SUMMARIZATION STRATEGIES

The major aim of this paper is to demonstrate that different summation strategies can result in different performance evaluations for local councils. It is clear that presenting data for all 152 councils in NSW would be both beyond the scope of a journal article, as well as superfluous to the main purpose of the paper (although it should be noted that these results are available from the corresponding author). However, it was also important that the study should not stand accused of selection bias. We thus elected to present the results from a stratified sample of data. To achieve our stratification, we first ranked all councils using factor analysis. We then selected the highest representative from each FSR as they appeared in the top third (upper band), center (mid band) and lower third (lower band) of the rankings. For instance, Temora was the highest ranked 'sound' council in the upper band, Oberon was the highest ranked 'sound' council in the mid band and Lithgow the highest ranked 'sound' council in the lower band. Only one 'strong' council had observations for all 10 performance indicators: it is thus the sole example of its category. This stratification is similar to the approach taken by Jacobs and Goddard (2007). The results - presented as the council ranking for ease of discussion - arising from the various summation strategies are detailed in Table 2.

⁶ The Renewals ratio seeks to measure whether a municipality is spending sufficient funds to renew assets relative to the deterioration in the asset base as estimated by the depreciation accruals.

Table 2
Comparison of Rankings Obtained Under Linear Methods, Unweighted

Council	FSR Rating	Standardization	Range Scaled Summation	Binary	Quartile Scoring	FA	PCA
Tumbarumba	Strong	10	10	1-8	1-4	22	18
Upper Band							
Temora	Sound	12	31	24-60	29-33	3	4
Kogarah	Moderate	1	1	61-95	13-19	1	1
Kyogle	Weak	36	35	96-121	48-56	13	38
Central Darling	Very Weak	129	131	122-130	83-92	118	28
Mid Band							
Oberon	Sound	50	68	61-95	35-47	66	15
Canterbury	Moderate	53	36	24-60	57-76	67	43
Byron	Weak	118	105	96-121	111-117	68	135
Gwydir	Very Weak	131	128	122-130	125-128	125	94
Lower Band							
Lithgow	Sound	6	8	24-60	93-100	110	124
Camden	Moderate	46	76	96-121	83-92	136	67
Carrathool	Weak	130	133	61-95	83-92	133	55
Greater Taree	Very Weak	136	136	131-135	136	130	114

It is clear from the results in Table 2 that the use of alternate unweighted linear summary methods produces substantially different rankings, which cannot be immediately reconciled with the FSR. For instance, Kogarah (which is ranked 'moderate' under FSR) ranked equal or higher than Temora (which received the higher FSR classification of 'sound') in all but one of the six alternate summaries. In addition, Kogarah ranked above Tumbarumba ('strong') in four of the

six alternative summary methods. Perhaps more perplexing was the result that Kyogle ('weak') ranked higher than Oberon ('sound') in half of the alternate compilations and higher or equal to Camden ('moderate') under every alternate linear summary method. Furthermore, Tumbarumba - the sole example of the highest 'strong' FSR - failed to rank above ten in four of the alternate performance indicator compilations.

This suggests that - in an unweighted context - the summary method selected has a significant effect on the rating which a local authority might be assigned. While some level of variation is to be expected, the degree of variation detailed is nonetheless extraordinary, particularly when one considers that the results presented are not 'cherry picked' exemplars, but rather the output of a stratified selection method which reflects the results obtained from the entire cohort. A particular instance of extraordinary variation is the case of Kogarah (with a FSR of 'moderate') which was ranked first under four of the compilation methods, but ranked 61-95 under the binary method. The explanation for this surprising result is found in our earlier observation that the validity of binary scoring rests on the soundness of designated benchmarks. Binary scoring was the only linear compilation method to utilize the TCorp (2013) ratio benchmarks. Thus, the variation observed for binary scoring reinforces the fact that it is absolutely critical for benchmarks to be set on the basis of sound reasoning.

Table 3 presents a correlation matrix which allows for a quick assessment of the level of similarity between the various compilation methods for the entire cohort. Not surprisingly, there is a good deal of similarity between the two methods employing versions of simple scaled summation (range scaled summation and standardization summation). The similarity between these two methods will be highest when data approximates a normal distribution and analysts wishing to employ simple summation algorithms should be guided by this observation. By way of contrast, binary scoring produced results distinctly dissimilar to the other methods employed and should probably only be used when analysts are confident in the benchmarks assigned to individual ratios. Quartile scoring exhibited lower levels of variation from the simple summation algorithms than did

binary scoring owing to the fact that it uses distribution attributes rather than exogenous benchmarks. Quartile scoring seems particularly suitable when the relative position of municipalities is the primary concern. FA and PCA produced similar rankings to one another, but dissimilar results to all other methods employed. This is not altogether surprising given that both methods strive to reduce the dimensionality of a data set. The fact that these summarization techniques produced rankings significantly different from all other methods could be attributed to the fact that the financial ratios employed by TCorp (2013) represent more than one dimension. Indeed, Drew and Dollery (2015) have recently demonstrated that the TCorp (2013) financial data reflect three major latent causal factors with orthogonal associations. Thus, the results arising from this study sound a note of caution for analysts interested in summarizing data into a single number for ranking purposes or simplicity: if the data reflect several distinct dimensions of performance, then there is a real risk that FA and PCA methods will be fundamentally flawed or at a minimum hide the strengths and weaknesses behind the single number.

Table 3
Pearson Correlation Matrix, Unweighted

	Standardization	Range Scaled Summation	Binary	Quartile Scoring	FA	PCA
Standardization	1.0000					
Range Scaled Summation	0.9698	1.0000				
Binary	0.5558	0.5958	1.0000			
Quartile Scoring	0.7329	0.7212	0.6949	1.0000		
FA	0.5910	0.4955	0.0715	0.2614	1.0000	
PCA	0.5900	0.4790	0.0742	0.2900	0.9557	1.0000

Table 4
Comparison of Rankings Obtained Under Linear Methods,
Weighted (Unweighted Rankings in Parentheses for
Comparative Purposes)

Council	FSR Rating	Standardi zation	Range Scaled Summa tion	Binary	Quartile Scoring	FA ^c	PCA ^c
Tumbarumba	Strong	9 (10)	17 (10)	9-16 (1-8)	3-7 (1-4)	22	18
Upper Band							
Temora	Sound	54 (12)	80 (31)	57-82 (24-60)	59-60 (29-33)	3	4
Kogarah	Moderate	1 (1)	2 (1)	57-82 (61-95)	15-15 (13-19)	1	1
Kyogle	Weak	52 (36)	57 (35)	108- 122 (96- 121)	69-73 (48-56)	13	38
Central Darling	Very Weak	128 (129)	134 (131)	127- 132 (122- 130)	116-119 (83-92)	118	28
Mid Band							
Oberon	Sound	75 (50)	84 (68)	90-106 (61-95)	69-73 (35-47)	66	15
Canterbury	Moderate	35 (53)	28 (36)	18-34 (24-60)	50-53 (57-76)	67	43
Byron	Weak	114 (118)	97 (105)	57-82 (96- 121)	106-110 (111-117)	68	135
Gwydir	Very Weak	129 (131)	128 (128)	127- 132 (122- 130)	128-128 (125-128)	125	94
Lower Band							
Lithgow	Sound	10 (6)	13 (8)	57-82 (24-60)	90-92 (93-100)	110	124
Camden	Moderate	72 (46)	95 (76)	108- 122 (96- 121)	99-103 (83-92)	136	67
Carrathool	Weak	135 (130)	135 (133)	90-106 (61-95)	114-117 (83-92)	133	55
Greater Taree	Very Weak	136 (136)	136 (136)	127- 133 (131- 135)	136 (136)	130	114

^c Dimensions can be weighted, but weighting all variables will not alter scores.

However, it appears that the FSR are weighted, although the exact method of compilation has not been detailed by TCorp (2013). Table 4 details the weighted results - using TCorp (2013) weights - for the same thirteen councils and comparison with the unweighted data immediately demonstrates the critical role that weights can play in any performance indicator summary. This is important to note given that (a) the result confirms the work of Jacobs and Goddard (2007) on the importance of weighting decisions and (b) it calls into question subjective weighting schemes which are not supported by theoretical insights or statistical methods, such as PCA or FA (Nardo et al. 2005). TCorp (2013) FSR weights appear to have been determined on a purely subjective basis (see Drew and Dollery 2014).

The weighting scheme imposed on the performance indicators (see Table 1) has resulted in the highest ranked council in terms of FSR (Tumbarumba) failing to rate as the highest municipality under any alternate linear summary. This represents a stark example of why FSR 'economic facts' cannot be considered in isolation from value judgements, weightings, trade-offs and the political dimension (Diefenbach 2009). A second example resides in Kogarah (moderate) which is now ranked equal or higher than Temora (sound) in each and every alternate method. Moreover, Kyogle ('weak') is now ranked higher or equal to Oberon ('sound') under four alternate linear summaries and continues to be ranked higher or equal to Camden ('moderate') under every alternative system. It is thus clear that weightings have an important effect on the rankings assigned to individual councils. Accordingly, it would appear reasonable for agencies employing weights for individual performance indicators to present both weighted and unweighted summaries so that information users can quickly assess the effect of weighting decisions.

Whilst weighting of performance indicators has changed the results for individual local authorities, it has done little to reduce the degree of incongruence between FSR and

alternate linear summaries for the entire cohort. Table 5 provides details of Pearson correlation coefficients for the six summaries. It is noteworthy that weighting has increased the degree of correlation for binary and quartile summation with standardization and range scaling. This seems to be an artifact of the situation wherein the highest weighted ratios had the most normal distributions along with medians very close to the respective benchmarks (see Table 1). Perhaps of greater interest is the fact that applying weights further reduced the degree of similarity between the two methods focusing on dimension reduction and the summation approaches (scaled and standardized). This is consistent with our earlier observation that the TCorp (2013) financial ratios reflect multiple dimensions. Moreover, the results suggest that the highest weighted ratios are associated with different dimensions.

In sum, the data presented for the various compilation strategies clearly demonstrate that the particular method selected for the ‘objectifications of the otherwise vague and abstract’ (Hopwood 1987) acts as a major determinant of the rating or ranking that a given municipality might receive. As we noted earlier, some variation in results arising from alternate summary algorithms is to be expected. However, the extraordinary degree of variation wherein a nominally ‘weak’ council scheduled for compulsory consolidation as a result of its assessment can rank above ‘moderate’ and ‘sound’ councils deemed to be financially sustainable should be of considerable concern to public policymakers. It is also an important result since the scholarly literature has largely neglected the matter of how much influence the summarization method (and associated weights) has on performance ratings. Section 5 examines the constitutive implications for two councils which received starkly contrasting results as an illustration of the types of behaviors which may be elicited by disparate ratings.

Table 5
Pearson Correlation Matrix, Weighted

	Standardization	Range Scaled Summation	Binary	Quartile Scoring	FA ^d	PCA ^d
Standardization	1.0000					
Scaled Summation	0.9667	1.0000				
Binary	0.6721	0.6905	1.0000			
Quartile Scoring	0.8470	0.8133	0.7555	1.0000		
FA*	0.3961	0.3232	0.0726	0.2479	1.0000	
PCA*	0.3657	0.2715	0.0526	0.2371	0.9557	1.0000

^d Dimensions can be weighted, but weighting all variables will not alter scores.

CONSTITUTIVE IMPLICATIONS

Section 2 discussed, in general terms, how ‘facts created by the craft give rise to an influential language and set of categories for conceiving and changing the organization’ (Hopwood 1987, 229). Section 5 takes specific examples drawn from different ends of the summary spectrum to explore the constitutive effects that might eventuate as a result of the choice of performance indicator summary method. We have clearly chosen the two councils from our stratified sample which have results of the greatest contrast. However, it should be borne in mind that (a) the sample is broadly representative of the entire cohort (for instance, Cooma-Monaro nominally classified as “weak” was also ranked well above “moderate” and “sound” councils under alternate summary methods) and (b) that section 5 simply strives to illustrate the types of behavior which may be promoted by different versions of ‘particular objectifications of the otherwise vague and abstract’. Table 6 details the FSR of our two case studies – Kyogle (rated ‘weak’ by TCorp (2013)) and Oberon (rated ‘sound’). The summary suggests that there is far less difference in the financial ratios for the two councils than may have been indicated by the TCorp (2013) ratings. For example, both councils fail to meet the benchmarks for the two highest weighted FSRs (Operating and Own Source ratios). Moreover, whilst Oberon had far better results for the Buildings and Infrastructure Renewal and Capital Expenditure ratios, Kyogle dominated in the

liquidity metrics (Cash Expense and Unrestricted Current ratios).

Table 6
Financial Ratios for Kyogle and Oberon Councils, 2011.

Ratio	Kyogle	Oberon
Operating	-11.2%	-6.9%
Own Source revenue	58.8%	52.0%
Unrestricted Current	7.52	3.30
Interest Cover	30.56	72.27
Infrastructure	0.21	0.15
Backlog		
Debt Service Cover	26.44	22.66
Capital Expenditure	0.65	1.47
Cash Expense	11.2	1.7
Buildings and	0.53	1.41
Infrastructure		
Renewal		
Asset Maintenance	0.91	0.82

Given Kyogle's rating of 'weak' (TCorp 2013), it seems reasonable to presume that senior management and elected representatives are likely to seek improvement in this categorization, not least because the 'weak' rating has been used to justify a recommendation for merger with adjoining municipalities. Perhaps the most obvious path to improving the 'weak' rating is to address the two highest weighted performance indicators: Operating ratio and Own Source Revenue ratio, which together account for over a third of the total FSR (see Table 1). Increasing fees, charges and property taxes will result in a positive response from both performance indicators. Accordingly, these revenue-raising measures, avoided by elected representatives due to potential backlash from the local community, may suddenly become visible to Kyogle management. In fact, at a recent meeting the Kyogle council voted to lift rates by 22% above the rate peg prescribed by the Independent Pricing and Regulatory Authority (Broome 2014). Moreover, the unflattering 'economic facts' of a 'weak' municipality disseminated to the local community may well provide a receptive context for revenue-raising efforts. Indeed, a recent survey stated that 78% of residents were opposed to a merger and that 48%

were willing to pay considerably higher council rates in order to improve the financial sustainability of the municipality (The Casino Times 2014).

This same 'objectification of the otherwise vague and abstract' (Hopwood 1987) may also serve as motivation for reduction in expenditure through reduced service quality, eliminating discretionary services (which are not measured as positive elements of FSR performance indicators) or making staff redundant. In fact, the most recent financial reports record a 10.16% drop in total expenditure (Kyogle Council 2014). This illustrates the types of responses which local councils may make in response to poor financial sustainability ratings. There is nothing intrinsically bad about the response in itself. We are simply interested in how different ratings derived from alternative summarization strategies may have caused municipal officials to act in different ways.

For instance, it is clear from Section 4 that Kyogle would be unlikely to receive the same categorization under five of the six alternate summary methods. Had FA been used to rate municipalities - whereby Kyogle was ranked 13th in the cohort - it is entirely possible (but by no means definite) that municipal officials may not have pursued such a significant increase to the tax rate. Moreover, it is reasonable to suggest that were it not for the 'weak' rating - and subsequent recommendation for consolidation - residents would not have been so willing to express support for higher council taxes.

At the opposite end of the TCorp (2013) financial sustainability rating spectrum, Oberon illustrates some of the organizational responses to a favorable financial sustainability assessment. Oberon was one of 32 municipalities which received the second highest categorization of 'sound' by TCorp (2013). In response to this assessment management may well be disinclined to aggressively reduce expenditure and be content with increases to local property taxes prescribed under the annual 'rate-cap' set by the NSW Independent Pricing and Regulatory Tribunal (IPART). Indeed, Oberon has not made any application for Special Rate Variation (to exceed the rate-cap) since the TCorp (2013) assessments and council documents yield no evidence to suggest that an increase in

property taxes has been entertained (IPART 2014). In addition, a 'sound' FSR provides little defense (on financial sustainability grounds) to any community demands for new services or improved quality of services. The 'economic facts' created by TCorp (2013) thus may make it more difficult for management, local residents or elected representatives to create a more efficient Oberon. For instance, recent financial statements detail a 6.8% increase in municipal expenditure against a 2.6% fall in council income for the same period (Oberon 2014). Yet under four of the alternative summary schemes Oberon performs well below the cohort median. Moreover, the fifth scheme (i.e. FA score) is a mere two points above the median.

Our brief case study illustrates some of the different behavioral responses which may be engendered by alternative 'objectification[s] of the otherwise vague and abstract' (Hopwood 1987). We have thus sought to demonstrate that decisions regarding the compilation method employed to reduce performance indicators to a single number not only affect the actual rating conferred on a council, but may also have a significant effect on future organizational behavior.

CONCLUSIONS

In this paper we have demonstrated that the choice of performance indicator summarization method is a major determinant of the rating conferred on a given municipality. Furthermore, by drawing on constitutive accounting theory we have illustrated that differences engendered by the various methods are far more than an academic curiosity: they possess the potential to alter the behavior of individual organizations.

A number of valuable lessons can be drawn from the empirical results. Firstly, it is clear from the evidence that choice of compilation method matters. This may seem a rather obvious conclusion, but the scholarly literature has largely neglected this question. Moreover, our analysis points to some important methodological considerations. For instance, we have demonstrated that binary scoring can lead to some distinctly disparate rankings owing to its reliance on performance benchmarks. This suggests that binary scoring

should only occur when analysts are certain that benchmarks are robust. We also observed that range scaled summation and standardized summation lead to similar results, but that range scaling should be preferred owing to its resistance to skewing. In addition, we noted the merit of quartile scoring, particularly in the case where analysts are primarily concerned with relative position.

Perhaps the greatest lesson associated with the analysis relates to the need for a careful consideration of the dimensions of the performance indicator set. Municipalities produce a heterogeneous mix of local services which provide *prima facie* evidence against the presumption of just a single dimension of performance (Drew and Dollery 2015). Accordingly, a performance indicator system based on financial data alone will struggle to holistically capture municipal sustainability. Moreover, the use of financial data poses problems given that it is necessarily orientated towards past performance. However, even within a performance indicator suite composed entirely from financial data it is entirely possible that a number of performance dimensions may exist. For instance, amongst the TCorp (2013) FSR we identified three major largely unrelated latent constructs. The existence of multiple dimensions means that FA and PCA compilation methods may be fundamentally flawed or - at a minimum - conceal strengths and weaknesses beneath the single number.

Our empirical analysis also demonstrated the significant effect which weighting has on the relative rankings of municipalities. This leads us to recommend greater levels of transparency where weights are applied, both in terms of justifying the individual weights, but also in disclosing the effect which weights had on the final ranking. In principle, justification would be based on empirical evidence or sound and clearly articulated policy arguments. Disclosure of the effect of weighting schemes could simply be made by providing information users with both the weighted and unweighted rankings.

There will always be a temptation for regulatory authorities to summarize suites of performance indicators into a single number to facilitate rankings, aid dissemination or simply

reduce the financial literacy demands imposed on users of the information. However, it is clear from our analysis that the compilation strategy itself can have a significant effect on the ratings assigned. It is thus critical that great care is taken in choosing a suitable compilation strategy and that the process is entirely transparent throughout, particularly in a 'high stakes' environment, such as the proposed NSW compulsory council consolidations.

REFERENCES

- Andrews, R., G. Boyne, J. Law and R. Walker (2005). 'External Constraints on Local Service Standards: The Case of Comprehensive Performance Assessment in English Local Government.' *Public Administration* 83(3): 639–656.
- Andrews, R., G. Boyne and R. Walker (2011). 'Impact of management on administrative and survey measures of organizational performance.' *Public Management Review* 13(2): 227-255.
- Bevan, G. and C. Hood (2006). 'What's Measured is What Matters: Targets and Gaming in the English Public Health Care System.' *Public Administration* 84(3): 517-538.
- Bird, S., D. Cox, V. Farewell, H. Goldstein, T. Holt and P. Smith (2005). 'Performance Indicators: Good, Bad, and Ugly.' *Journal of Royal Statistical Society* 168(1): 1-27.
- Broome, H. (2014). 'Kyogle Votes to Lift Rates 22% Above Peg Over 20 Years,' *Northern Star*. Lismore: Northern Star.
- Brown, K. (1993). 'The 10-Point Test of Financial Condition: Toward an Easy-To-Use Assessment Tool for Smaller Governments.' *Government Finance Review* 9(1): 21-25.
- Bouckaert, G. and J. Halligan (2008). *Managing Performance: International Comparisons*. London: Routledge.
- Casino Times (2014). 'Let's NOT Talk Turkey,' *Casino Times*. Casino: Casino Times.

Congressional Budget Office (1978). *City Need and the Responsiveness of Federal Grants Programs*. Washington D.C.: Government Printing Office.

Cunningham, G. and J. Harris (2005). 'Toward a Theory of Performance Reporting to Achieve Public Sector Accountability: A Field Study.' *Public Budgeting & Finance* Summer 2005: 15-42.

Department of Local Government (2006). *Standard Contract of Employment*. Sydney: Department of Local Government.

Diefenbach, T. (2009). 'New Public Management in Public Sector Organisations: The Dark Sides of Managerialistic Enlightenment.' *Public Administration* 87(4): 892-909.

Division of Local Government. (2013). *Comparative Information on NSW Local Government – Measuring Local Government Performance*. Sydney: Division of Local Government.

Drew, J. and B. Dollery (2014). 'Estimating the Impact of the Proposed Greater Sydney Metropolitan Amalgamations on Municipal Financial Sustainability.' *Public Money & Management* 34(4): 281-288.

Drew, J., M. Kortt and B. Dollery (2013). 'Did the Big Stick Work? An Empirical Assessment of Scale Economies and the Queensland Forced Amalgamation Program.' *Local Government Studies*: DOI: 10.1080/03003930.2013.874341.

Drew, J. and B. Dollery (2015). 'A Factor Analytic Assessment of Financial Sustainability: The Case of New South Wales Local Government.' *Australian Accounting Review*, In Print.

Dunteman, G. (1989). *Principal Components Analysis*. California: Sage Publications.

Game, C. (2006). 'Comprehensive Performance Assessment in English Local Government.' *International Journal of Productivity and Performance Management* 55(6): 466 – 479.

Hawke, L. (2012). 'Australian Public Sector Performance Management: Success or Stagnation?' *International Journal of Productivity and Performance Management* 61(3): 310 – 328.

Hendrick, R. (2004). 'Assessing and Measuring The Fiscal Health of Local Government – Focus on Chicago Suburban Municipalities.' *Urban Affairs Review* 40(1): 78-114.

Hood, C. (2007). 'Public Service Management by Numbers: Why Does it Vary? Where Has it Come From? What are the Gaps and the Puzzles?' *Public Money & Management* 27(2): 95-102.

Hopwood, A. (1987). 'The Archaeology of Accounting Systems.' *Accounting, Organisations and Society* 112(3): 207-234.

Independent Local Government Review Panel (ILGRP). (2013a). *Future Directions for NSW Local Government - Twenty Essential Steps*. Sydney: ILGRP.

Independent Local Government Review Panel (ILGRP). (2013b). *Revitalising Local Government*. Sydney: ILGRP.

Independent Pricing and Regulatory Tribunal (2014). *Applications and Determinations on Special Rate Variation Submissions*. Sydney: IPART.

Jacobs, R. and M. Goddard (2007). 'How Do performance Indicators Add Up? An Examination of Composite Indicators in Public Services.' *Public Money & Management* 27(2): 103-110.

Kelly, J. and D. Swindell. (2002). 'A Multiple Indicator Approach to Municipal Service Evaluation: Correlating Performance Measurement and Citizen Satisfaction Across Jurisdictions.' *Public Administration Review* 62(5): 610-621.

Kimm J. and C. Mueller (1978), *Factor Analysis – Statistical Methods and Practical Issues*. California: Sage Publications.

Kloot, L. (1999). 'Performance Measurement and Accountability in Victorian Local Government'.

International Journal of Public Sector Management 12(7): 565-583.

Kloot, L. and J. Martin (2000). 'Strategic Performance Management: A Balanced Approach to Performance Management Issues in Local Government'. *Management Accounting Research* 11: 231-251.

Kloha, P., C. Weissert and R. Kleine (2005). 'Developing and Testing a Composite Model to Predict Local Fiscal Distress'. *Public Administration Review* 65(3): 313-323.

Kyogle Council (2014). *Audited Financial Statements 2014*. Kyogle: Kyogle Council.

Maher, C. and S. Deller (2013). 'Measuring the Impacts of TELs on Municipal Financial Conditions', in: Levine, H., B. Justice, and E. Scorsone (Eds), *Handbook of Local Government Fiscal Health*. Burlington MA: Jones and Bartlett Learning.

Moynihan, D. P. (2008). *The Dynamics of Performance Management: Constructing Information and Reform*. Washington D.C.: Georgetown University Press.

Nardo, M., M. Saisana, A. Saltelli and S. Tarantola (2005). *Tools For Composite Indicator Building*. Italy: European Commission Directorate-General Joint Research Centre.

Nathan, R. and C. Adams (1976). 'Understanding Central City Hardship.' *Political Science Quarterly* 91(1): 47-62.

Oberon Council (2014). *Audited Financial Statements 2014*. Oberon: Oberon Council.

Pollitt, C. and G. Bouckaert (2000). *Public Management Reform: A Comparative Analysis*. Oxford: Oxford University Press.

Queensland Treasury Corporation (QTC). (2008). *Financial Sustainability in Queensland Local Government*. Brisbane: QTC.

Saisana, M., A. Saltellia and S. Tarantola (2005). 'Uncertainty and Sensitivity Analysis Techniques as Tools

for Quality Assessment of Composite Indicators.’ *Journal of Royal Statistical Society* 168(2): 307-323.

Saltelli, A. (2007). ‘Composite Indicators: Between Analysis and Advocacy.’ *Social Indicators Research* 81: 65-77.

TCorp (2013). *Financial Sustainability of the New South Wales Local Government Sector*. Sydney: TCorp.

U.S. Department of the Treasury, Office of State and Local Finance (1978). *Report on the Fiscal Impact of the Economic Stimulus Package on 48 Large Urban Government*. Washington D.C.: Government Printing Office.

Van Dooren, W. and S. van de Walle (eds) (2008). *Performance Information in the Public Sector*. London: Palgrave Macmillan.

Van Dooren, W., G. Bouckaert and J. Halligan (2010). *Management*. London: Routledge.

Van Thiel, S. and F. Leeuw (2002). ‘The Performance Paradox in the Public Sector.’ *Public Performance and Management Review* 25(3): 267–281.

Zaleznik, A. (1989). *The Managerial Mystique – Restoring Leadership in Business*. New York: Harper and Row.

Zafra-Gomez, J., A. Lopez-Hernandez and A. Hernandez-Bastida (2009). ‘Evaluating Financial Performance in Local Government: Maximising the Benchmarking Value.’ *International Review of Administrative Sciences* 75(1): 151-167.

Zygmunt, C. and M. Smith (2014). ‘Robust Factor Analysis in the Presence of Normality Violations, Missing Data, and Outliers: Empirical Questions and Possible Solutions.’ *Quantitative Methods for Psychology* 10(1): 40-55.