

PERFORMANCE MEASUREMENT SYSTEM EFFECTIVENESS IN PUBLIC SECTOR IT OUTSOURCING CONTRACTS

Brian Perrin and Graham Pervan
Curtin University of Technology, Perth, Western Australia

Abstract

A key issue in the successful management of an Information Technology (IT) outsourcing contract is the measurement of the vendor's performance to ensure that all agreed outcomes are achieved. An effective performance measurement system (PMS) provides both control and accountability of the vendor and the contract arrangements. Despite the recognised importance of performance measurement to the overall success of IT outsourcing, a review of the literature highlights that research in this area is very limited. Based on a survey of large public sector organisations and universities throughout Australia, this paper develops and explores a three-dimensional model initially developed by Bouckaert (1993) and later refined by Streib and Poister (1999) to measure the effectiveness of performance measurement systems used in public sector IT outsourcing contracts. This research supports Bouckaert's (1993) proposition that an effective performance measurement system needs to include three essential characteristics - be valid, legitimate and functional. This research contributes to the existing literature on the choice and implementation of PMSs used in public sector IT outsourcing service contracts, and provides some practical guidance to public sector managers on the design of effective PMSs. The results also highlight that an effective PMS is a dynamic process that is not static and should be continually improved.

Keywords: performance measurement systems effectiveness, public sector IT outsourcing, validity, functionality, legitimacy

Introduction

A key issue in the successful management of an Information Technology (IT) outsourcing contract is the measurement of the vendor's performance to ensure that all agreed outcomes are achieved. An effective performance measurement system (PMS) provides both control and accountability of the vendor and the contract arrangements. The precise requirements and extent of the performance measurement system will vary from contract to contract and will depend upon the nature of the work, the type of contract, transaction characteristics and the relationship between the vendor and the agency (ANAO, 2001; SSC, 2002). Performance measurement systems can range from the most basic acceptance of a delivery and payment, to more complex situations that require extensive involvement of both client and vendor staff throughout the contract term. The responsibility for measuring performance can be either 'direct' or 'devolved', or a mix of both (JCPAA, 2000). With direct measurement, the agency conducts most of the measurement, whereas with devolution the vendor is largely responsible.

A performance measurement system is an integral part of the control system within an organisation. A performance measurement and control system in a business is like the controls in a car. 'The steering, accelerator, and brakes allow the driver to control direction and speed; instrumentation on the dashboard provides critical information about actual speed and early warning about potential problems with the cars key operating system' (Simons, 2000, p. 5).

Despite the recognised importance of performance measurement to the overall success of IT outsourcing, a review of the literature highlights that research on measuring the effectiveness of performance measurement systems used in public sector IT outsourcing contracts is very limited. An effective performance measurement system is an essential part of the planning and control process that allows firms to communicate strategy, track performance against targets, take corrective action when required, assess the quality of services provided and be used to help guide future strategies (Langfield-Smith and Smith, 2004). Further it is claimed that an effective performance measurement system is a tool or aid which helps a business to successfully manage the contract, enhance accountability, provide confidence that the outsourcing relationship is being effectively managed and is achieving all of its objectives (Kern and Willcocks, 2000a; Langfield-Smith and Smith, 2004). The consequences of an ineffective performance measurement system include unnecessary delays in resolving problems, additional costs, poor quality service and damage to the ongoing relationship. This paper develops and explores a three-dimensional model initially developed by Bouckaert (1993) and later refined by Streib and Poister (1999) to measure the effectiveness of performance measurement systems used in public sector IT outsourcing contracts. This research supports Bouckaert's (1993) proposition that an effective performance measurement system needs to be valid, legitimate and functional. The remainder of this paper is divided into five sections: (1) Literature Review, (2) Model and Propositions, (3) Research Method, (4) Results and (5) Summary, Limitations and Future Research.

Literature review

Over the last ten years a considerable amount of research on IT outsourcing has been conducted. This research has focused on numerous areas including: determinants of outsourcing, reasons for and against outsourcing, types and extent of outsourcing, identification of best practices and operational issues, measures of outsourcing success, risk associated with outsourcing relationship issues. However, very little research evidence has been accumulated on the management of IT outsourcing contracts or on identifying the factors that influence the design and effectiveness of performance measurement systems used in public sector IT outsourcing contracts.

A number of research studies refer to the importance of performance measurement systems. However, very few examined the effectiveness of performance measurement systems used in IT outsourcing arrangements. An effective performance measurement system is critical to the success of the IT relationship (McFarlan and Nolan, 1995; Willcocks, Lacity and Fitzgerald, 1995) because it helps a business to successfully manage the contract, enhance accountability and provide confidence to the organisation that the outsourcing relationship is being effectively managed and is achieving all of its objectives.

In recent years, research on performance measurement systems has mainly focused on the different types of systems that are currently used in organisations. For example, new contemporary performance systems such as the balanced scorecard, results, and determinants

framework, and the performance pyramid (Cavalluzzo and Ittner, 2004; Ittner and Larcker, 1998; Kaplan and Norton, 1997; Poister and Streib, 1999) are currently used. Australian studies that refer to performance methods used in outsourcing include Domberger, Hensher and Wedde (1993), Hall and Rimmer (1994), Dean and Kiu (2001) and Dean (2000). These studies found that both private and public sector organisations made little use of performance indicators to control or assess the vendor's performance. In cleaning contracts, private sector organisations relied mainly on inspection, whereas public sector agencies used certification (Domberger, Hensher and Wedde, 1993). Most contracts were based on a fixed price and functional specification. Dean and Kiu's (2001, p. 1) 'key finding was that organisations rely on inspections by their own employees or contractor checklists, but that these practices are in conflict with their views on best practices'. Dean's (2000) research considers issues inherent in the measurement and monitoring of quality in the context of contracted services. 'Key findings include the lack of consensus about quality definitions and measures, the lack of guidance from the literature in terms of performance monitoring, and the need to ensure that the costs of quality measurement activities do not exceed the benefits of contracting' (Dean, 2000, p. 1).

Research in the area also indicates that there are a number of different approaches to performance measurement systems that are currently used in IT outsourcing contracts (Industry Commission, 1996). For example, some systems are very complex and include a broad range of service levels, targets, key performance indicators, benchmarking techniques, customer satisfaction surveys, and a detailed reporting / review process. Others are very simple and rely solely on a method of direct inspection (Dean and Kiu, 2001; Hall and Rimmer, 1994; Halvey and Melby, 2000; IC., 1996; JCPAA., 2000; MAB/MIAC, 1997; SSC., 1997). As mentioned previously, in some systems the performance reports are prepared internally by the host agency or alternatively are prepared by the vendor or a third party.

Measuring the design and effectiveness of the performance measurement systems used in public sector IT outsourcing service contracts is an area that has attracted very little research attention. In general, research has mainly been descriptive studies that have focused on the design, types and implementation issues. Very little research has focused on the effectiveness or usefulness of performance measurement systems in either the private or public sectors. For example, there is no specific research that relates the effectiveness of performance measurement systems to the success of IT outsourcing contracts. However, there are a number of research studies that refer indirectly to the importance of performance measurement systems in relation to the overall success of the outsourcing arrangement (Kern and Willcocks, 2000a; McFarlan and Nolan, 1995; Willcocks, Lacity and Fitzgerald, 1995). Willcocks et al. (1995) claim that an effective PMS can critically affect the overall success or failure of the IT outsourcing contract. Further, they state the contract is the fundamental building block of an effective PMS. Organisations found it difficult to evaluate vendor bids because of the lack of prior PMS before outsourcing (Lacity, Willcocks and Feeny, 1995; Willcocks, Lacity and Fitzgerald, 1995).

The design of performance measurement systems (PMS) in IT outsourcing contracts is a complex process due to the intangibility and heterogeneity characteristics of the IT service provided. The appropriate design, implementation and use of the PMS is of critical importance to the success of the outsourcing contract (Lohman, Fortuin and Wouters, 2003; Neely and Bourne, 2000; Willcocks, Lacity and Fitzgerald, 1995). The PMS should be useful and a benefit to the organisation and therefore should not be too technical or detailed. The measurement systems need to be kept simple, meaningful and only measure those areas that

need to be measured (Halvey and Melby, 2000; Willcocks, Lacity and Fitzgerald, 1995). The design of the PMS should take into account the interests and expectations of all stakeholders involved in the outsourcing process (Kern and Willcocks, 2000b). Research on PMS supports this concept of a 'balanced' and 'integrated' approach to the design of performance measurement systems used in both the private and public sectors. A poorly designed PMS may not only be of no benefit to the organisation but also may result in dysfunctional behaviours (Bouckaert, 1993; Brignall and Modell, 2000).

The model and propositions

In this research, the instrument that was used to measure the effectiveness of the performance measurement system used in public sector IT outsourcing service contracts is based on the three-dimensional model initially developed by Bouckaert (2003) and later refined by Streib and Poister (1993). Both authors' research focused on the effectiveness of performance measurement systems used in US municipal government environments, which is quite different to the PMS used in Australian public sector IT outsourcing service contracts. However, the basic principle or framework upon which the three-dimensional model is founded is still appropriate to this research and therefore can still be effectively applied to the Australian public sector environment. Bouckaert's model was modified to suit the Australian environment and the specific characteristics of public sector IT outsourcing service contracts.

An effective performance measurement system should include more than just technical characteristics and should also take into account all three dimensions: validity, legitimacy and functionality (Streib and Poister, 1999). A brief description of Bouckaert's (1993) and Streib and Poister's (1993) interpretation of these three dimensions is outlined in the following sections:

Validity

Validity refers to the traditional technical requirements such as accuracy, reliability, timeliness, dependability, transparency and measurability. 'A valid measure is one that is sound, cogent, convincing and telling' (Streib and Poister, 1999, p. 31). A valid measure can also be described as one that is well grounded, relevant, meaningful and actually measures the characteristics it claims to measure (Bouckaert, 1993). Streib and Poister (1997) modified Bouckaert's measure of validity by adding other items such as best practice principles, desirability of measures, quality of measures, understandability and the need for clear objectives.

Legitimacy

Legitimacy refers broadly to the principle of controllability and acceptance of the measures used in the performance measurement system. An effective PMS is one that includes measures that are acceptable to all the stakeholders involved in the IT outsourcing contract and are not imposed on them by external sources such as funding organisations (Streib and Poister, 1999). There are many different stakeholders that exist within both client and vendor organisations that are involved in the IT outsourcing activity. These include: senior managers, business managers, IT managers, IT staff, end users and external fund providers (Poister and Streib, 1999). Each of these stakeholders has different perceptions and expectations of IT outsourcing and their interests should be taken into account when designing the PMS. To enhance accountability and controllability the measures need to be developed and approved

by those parties that are responsible for them and can take corrective action when necessary (Lacity and Willcocks, 2001). The relationship between legitimacy and effective performance measurement systems proposed by Bouckaert (1993) is supported by institutional theorists who claim that government organisations that conform to external expectations make little use of these systems for internal purposes (Bouckaert, 1993; Geiger and Ittner, 1996). In essence the users do not have any ownership in the system that is imposed from above.

Functionality

Functionality refers to the purpose or usefulness of the performance measurement system to the organisation. An effective performance measurement system should provide the organisation with usable data and have clear identifiable benefits (Lawton, McKevitt and Millar, 2000). An effective performance measurement system is one that provides relevant and useful information that assists managers in the decision-making process and helps maintain accountability. 'Performance measures are simply a means to an end that add no value to an organisation until someone takes action on the basis of the measures and the insight they contain' (Streib and Poister, 1999, p. 45). When designing performance measurement systems organisations need to clearly define the purpose of the performance measures and ensure they are correctly aligned with the organisations objectives and strategies. A poorly designed performance measurement system may not only be of no benefit to the organisation but also result in dysfunctional behaviours (Neely, Bourne, and Kennerley, 2003). An example of this dysfunctional behaviour is seen in the on-time train performance indicator that is used to measure the efficiency and effectiveness of passenger trains. In this example, train drivers maintain the target of on-time train running by not stopping at all of the required stations.

In summary, Bouckaert (1993) determines the effectiveness of a performance measurement system by measuring the three conditions – validity, legitimacy and functionality. Bouckaert (1993) develops eight (8) levels of effectiveness. A system that has high levels of each condition would be considered the 'optimum system'. A system that has three low levels would be classified as the 'worst case' (Bouckaert, 1993). Based on a similar framework to the model developed by Bouckaert (1993), Streib and Poister (1993) surveyed large municipal government organisations in the US to determine the usefulness of their performance measurement systems. Similar to Bouckaert, they measured validity, legitimacy and functionality (Streib and Poister, 1999). They also classified the systems examined in the survey into eight (8) classifications. The finding from Streib and Poister's (1999) research identified a number of strengths and weaknesses in the conditions that were examined and provided excellent support for the original model proposed by Bouckaert. For example, in the validity area 68% of the respondents indicated they had trouble measuring quality and that staff lacked the analytical skills needed to effectively analyse the performance measurement data. Respondent answers to the functionality item indicated that their performance measurement system enhanced accountability and improved the quality of decision making (Streib and Poister, 1999).

Based on the above discussion it is proposed in this research that the design of an effective performance measurement system used in public sector IT outsourcing contracts should be balanced and include all three dimensions proposed by Bouckaert (1993). An effective performance measurement system needs to be valid, legitimate and functional. This is summarised in the following Figure 1.1 and propositions:

Performance Measurement System Effectiveness

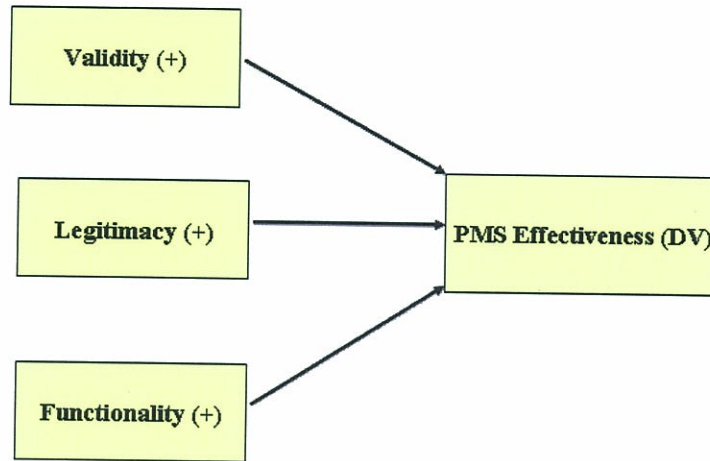


Figure 1.1: Conceptual Model

Proposition 1 (P1):

An effective performance measurement system used in public sector IT outsourcing contracts is positively influenced by the level of validity included in the design.

Proposition 2 (P2):

An effective performance measurement system used in public sector IT outsourcing contracts is positively influenced by the level of legitimacy included in the design.

Proposition 3 (P3):

An effective performance measurement system used in public sector IT outsourcing contracts is positively influenced by the level of functionality included in the design.

As highlighted in the above conceptual model, all three characteristics (i.e. validity, legitimacy and functionality) need to be equally represented in the ultimate design of an effective performance measurement system. For example, an effective performance measurement system is one that is relevant, useful; acceptable to all stakeholders and satisfies all the technical requirements. In summary, it is proposed in this research that the concepts highlighted in Bouckaert's (1999) model can be used to help design and effective performance measurement system that can be used to help manage the IT outsourcing arrangement.

Research method

The information outlined in this paper was collected via a survey that was mailed to 323 large public sector organisations and universities throughout Australia. The public sector was further divided into federal, state and local government organisations. Selection of the target population for this study was based on theoretical sampling as recommended by Neuman (1993). In this study, the criterion that was used to select the survey population was large public sector organisations that could be involved in IT outsourcing. Large public sector organisations were targeted because it was considered they would be more likely to have

outsourced an IT activity than a small organisation that had limited staff and computers. It was also presumed that gaining access to confidential information on IT contracts would be easier to obtain in the public sector than the private sector. Based on the information available, federal and state government organisations were classified as large if they employed over 500 staff. Local government organisations were classified as large if the population coverage for the shire or council was over 50,000 residents.

Respondent details needed for the survey mailing list was initially obtained from the Internet and purchased Australian Government Directories. Details of each target respondent (i.e. name, job title, address, state, telephone contact) were checked by telephoning either the Chief Information Officer (CIO) or the contact person nominated on the initial list collected. Each telephone call to the contact person, also verified the organisation address and details of the person in charge of the IT function within the organisation.

In developing the questionnaire and instrument, a number of best practice procedures were followed. For example, procedures proposed by Dillman (2000), Neuman (2000) and Andrews (2000) were used to enhance the quality of data, increase reliability and validity and to improve the response rate. In this study, multi-item questions were used to measure each of the variables outlined in the research model. Roberts (1984, p. 63) claims that multi-item questions are superior to single item scales because many constructs 'are complex concepts and this complexity is better covered if more than one question is asked'. Further, Roberts (1999) states that multi-item questions increase the level of validity and reliability. By asking multiple questions on the one construct 'minimises the risk that the wording substantially affects the answer' and also avoids misunderstanding possibly associated with the reliance on single item questions (Roberts, 1999, p. 63). This view is supported by Andrews (1999, p. 430) who states that 'as the number of answer categories goes up, data quality goes up – i.e., validity tends to increase and residual error tends to decrease'.

As detailed, the instrument that was used to measure the effectiveness of performance measurement systems used in public sector IT outsourcing contracts is based on the three dimensional model initially developed by Bouckaert (1984) and later refined by Streib and Poister (1993). A number of wording changes was made to Bouckaert's model to suit the Australian public sector and IT outsourcing environments. In total, twenty-seven questions in the survey instrument referred to PMS effectiveness. These questions were further divided into three (3) groups of questions: validity (10 items), legitimacy (8 items) and functionality (9 items). Full details of the questions used to measure validity, legitimacy and functionality is shown at Appendix 1. All three blocks of questions were measured on a seven-point Likert scale that was anchored at the end points only. Dillman (1999) states that by only labelling the end categories enhances validity of the data by making the choice for respondent easier to read and comprehend. In this study closed type questions, where the respondent is given a fixed response to choose, were selected because of the many advantages associated with a mail survey. These advantages include: easier and faster for respondents to answer, easier to code and statistically analyse and easier to compare answers from different respondents (Dillman, 2000). The questions were designed and worded to ensure they were clear and unambiguous to the respondent. To enhance the reliability and validity of the data collected, a pre-test of the instrument was conducted and a number of minor alterations were made to the wording of some questions to help clarify their meaning.

The questionnaire, covering letter and self-addressed postage paid envelope was mailed to the person in charge of the IT function within the organisation. This person was selected because

it was assumed they would have been involved in the IT outsourcing contract and the development of the performance measurement system used to manage the contract.

Results

In total, the survey questionnaire and covering letter was mailed to 323 large public sector organisations and public universities throughout Australia. Table 1 below shows how the target population for this study was divided into federal, state and local government organisations and the frequencies and percentages of the 'target population' in each category. For example, 44.3% and 37.8% were mailed to state and local government organisations respectively.

Category	Target Population	Target Population	Total Received		Response Rate	IT Outsourcing	Response Rate
	Number	%	Number	%	%	Yes	%
			(Q1 to 6)	(Q1 to 6)	(Q1 to 6)	(Q1 to 23)	(Q1 to 23)
Federal	22	6.8%	11	7.8%	50.0%	9	40.9%
State	143	44.3%	53	37.6%	37.1%	36	25.2%
Local	122	37.8%	62	44.0%	50.8%	25	20.5%
University	36	11.1%	15	10.6%	41.7%	3	8.3%
Total	323	100%	141	100%	43.7%	73	22.6%

Table 1: Survey Response Rate

Based on the target population, the response rate was 43.7% or 141 respondent organisations. The response rate for organisations which nominated they had outsourced an IT activity and completed the full questionnaire (questions 1 to 23) was 22.6% or 73 respondent organisations. Individual response rates for each level of government and universities are shown in Table 1 above. A response rate of 22.6% is similar to the response rate achieved by Cullen et al. (2000) in a recent survey of IT outsourcing practices in Australia. When compared to other IT surveys, a 22.6% response rate would be classified as high. For example, Lacity and Willcocks (2001) and Rouse (2000) only achieved a 17% response rate in their US and UK surveys of IT practices.

A Mann-Whitney U-test for non-response bias based on the degree of association between the first mail-out and the second mail-out was conducted. Overall, the number of respondents for the first mail-out was 72 (50.4%) and the second mail-out was 69 (49.6%). A comparison of both samples was made on key descriptive variables and research model variables. It was found, there is very little statistical difference between the first and second mail-outs. The z-score for each variable were below the critical level and that the significance tests are all well above the recommended level of .05. These results indicate that the sample is unbiased and is highly representative of the population. A chi-squared (goodness of fit) test was also conducted to determine whether the frequency of respondents to the survey was representative of the target population. The chi-squared (goodness of fit) test results for the frequency distribution of actual compared to expected level of respondents by category (federal, state, local and universities) is well below the recommended χ^2 critical value of 7.814 and above the significance level of .05. A chi-squared result of χ^2 of 3.094 and p of 0.377 would indicate that the sample is highly representative of the target population.

Table 1 above shows the highest number of respondents were from local government (44.0%) followed by state government (37.6%). The balance was divided between federal government

(7.8%) and universities (10.6%). Table 2 below details the frequency distribution of respondents by level of government and location.

Category	State								Total
	WA	SA	NSW	VIC	TAS	NT	QLD	ACT	
Federal	0	0	0	0	0	0	0	11	11
State	21	6	7	4	1	2	9	3	53
Local	9	4	20	13	1	1	14	0	62
University	2	0	6	3	0	0	3	1	15
Total	32	10	33	20	2	3	26	15	141
Total %	22.7%	7.1%	23.4%	14.2%	1.4%	2.1%	18.5%	10.6%	100%

Table 2 Frequency Distribution by Level of Government

Analysis of the geographic distribution of respondents shows the greatest number of respondents is from New South Wales (23.4%) followed closely by Western Australia (22.7%) and Queensland (18.5%). Further analysis of the actual compared with expected respondents indicates the state distribution of respondents is close to what was expected, only with the exception of Western Australia which was slightly above. This exception would be considered acceptable given that it was a Western Australian survey.

IT budget expenditure ranged from a minimum of \$0.250 million to a maximum of \$200 million. Overall, the average annual IT budget expenditure was \$19.03 million. Further analysis and grouping of IT expenditure budgets shows that 54.7% of the sample have budgets less than \$5 million and 72.3% have budgets less than \$15 million.

Based on the sample of 69 organisations that completed the full questionnaire, it can be seen in Table 3 below that 40.6% of the respondents outsourced less than 20% of their annual IT expenditure budget, 53.6% outsourced from 20% to 79% and only 5.8% outsourced more than 80%. Table 3 also shows that the percentage of IT expenditure budget outsourced ranged from a minimum of 0.63% to a maximum of 90%. Overall, the average extent of IT outsourcing of the 69 organisations was 27.9%.

	N	Minimum	Maximum	Mean	Std. Dev.
% of IT Budget Outsourced	69	0.63%	90.0%	27.9%	22.6%
Grouping %	0% to 19%	20% to 39%	40% to 59%	60% to 79%	80% to 100%
Frequency (n=69)	28	21	12	4	4
Frequency %	40.6%	30.4%	17.4%	5.8%	5.8%

Table 3 Extent of IT Outsourcing

The extent of IT outsourcing decisions can be broadly classified into three main types (Rouse, 2002). When organisations outsource less than 20% of their IT expenditure budget, the sourcing decision is classified as '*substantially insourced*'. For those organisations whose expenditure exceeds 80%, it is classified as '*substantially outsourced*'. The final type is classified as '*selectively outsourced*', for those organisations whose expenditure is greater than 20% and less than 80% of the IT budget (Lacity and Willcocks, 2000). Table 4 below

provides a comparison of the survey results of this research with recent Australian, UK and USA research findings:

Category	% of IT Budget Outsourced	Survey Results 2005	Lacity and Willcocks, 2000	Lacity and Willcocks, 2000	Cullen et al. 2001
		Aust.	UK	USA	Aust.
Substantially insourced	less than 20%	40.6%	30.0%	8.0%	47.0%
Selectively outsourced	20% to 79%	53.6%	65.0%	84.0%	46.0%
Substantially outsourced	80% or more	5.8%	5.0%	8.0%	7.0%
	Average	27.9%			28.0%

Table 4 Comparison of Extent of IT Outsourcing Surveys

As shown in Table 4 above, 53.6% of Australian public sector organisations are selectively outsourcing their IT budget expenditure. Only 5.8% of the sample is substantially outsourcing more than 80% of the IT expenditure budget. Lacity and Willcocks (2001) claim the low level of substantial outsourcing adopted by organisations in all countries may be due to its limited success. For example, substantial outsourcing contracts have experienced a 37% failure rate over the past ten years (Lacity and Willcocks, 2000). Table 4 above also shows that when compared to Cullen et al. (2000) the level of selective outsourcing has increased in Australia., from 46% to 53.6% and the substantially insourcing group has reduced from 47% to 40.6%. The 7% increase in selective outsourcing indicates that the market in Australia is maturing and organisations are growing in confidence. It also indicates that Australian public sector organisations are choosing to employ multiple vendors rather than rely on one primary vendor. This strategy provides a greater level of flexibility and reduces the level of risk.

Although there were 73 organisations that nominated they had outsourced an IT activity, only 71 respondents completed the questions that referred to value and duration of the contract. To overcome the problem of varying answers associated with multiple contracts, the survey asked organisations to refer to their largest IT outsourcing contract only. The contract values range from a minimum of \$0.015 million per annum to a maximum of \$234 million per annum. Overall, the average contract value was \$8.73 million per annum with a standard deviation of \$29.7 million. Further analysis in this area shows that 71.8% of the sample had contract values at less than \$5 million per annum and only 10% of all contracts were greater than \$15 million per annum.

The results show the initial contract duration ranged from a minimum of 1 year to a maximum of 10 years and an average of 3.8 years. Respondents also indicated that they extend their initial contracts by approximately 2.1 years taking the full term of the contract to an average 5.9 years. Analysis of the results also shows that 60% of the respondent organisations prefer contracts that are 3 years or less and that 86% prefer contracts that are 5 years or less. Only 14% of the organisations indicated they had contracts that lasted for 6 to 10 years. A shorter contract is the preferred option because it provides more flexibility by allowing the client to renegotiate the contract more frequently (S. K. Cullen et al., 2001). These results are similar to other research in this area. For example, Cullen et al. (1993) found that 79% of the contract duration was less than 3 years and 7% had contract terms of 6 to 10 years.

Respondents' indicated that the most frequently used pricing style was a fixed price (76.7%), followed by time and materials (63.0%) and cost plus (21.9%). An open book style was the least used (15.1%). The results are similar to other research in this area. For example, Currie (2001) found that both public and private sector organisations prefer to use fixed price contracts and fixed term contracts as opposed to open ended cost plus contracts.

Respondents to the survey indicated that the most frequently outsourced activities were: hardware support and maintenance 49%, application support and maintenance 46% and operations and facilities management 46%. The results also show that the least outsourced activities were business process re-engineering 4%, IT strategic planning 6% and education and training 10%. This survey supports the claim made by Cullen et al. (1996, p. 7) that organisations tend to outsource 'low risk mature stable activities, for which they could write detailed contracts and evaluation metrics'. The results of this survey are similar in most areas to the research findings of Cullen et al. (2001) and Lacity and Willcocks (2001).

As detailed, the instrument that was used to measure the effectiveness of performance measurement systems used in public sector IT outsourcing contracts is based on a three dimensional model initially developed by Bouckaert (2000) and later refined by Streib and Poister (1993). As shown in tables 5 to 7 below, twenty-seven items referred to PMS effectiveness. These questions were further divided into three groups of questions: validity (10 items), legitimacy (8 items) and functionality (9 items).

Validity	Not at All						Commonly	Mean	Std. Dev.
	1	2	3	4	5	6	7		
(a) We compare our PMS / SLA against similar programs or services.	28%	18%	6%	8%	25%	14%	1%	3.32	1.95
(b) Our PMS / SLA are easy to understand and are not too technical	8%	6%	14%	11%	36%	24%	1%	4.38	1.56
(c) When developing performance measures, we focus on what is important to measure rather than the availability of data	3%	6%	15%	21%	30%	24%	1%	4.46	1.35
(d) Our measures are derived from the mission, goals, objectives, and service standards established for our IT outsourcing contract	3%	14%	15%	24%	25%	17%	3%	4.15	1.46
(e) We do not have trouble measuring the quality of our programs and services	3%	15%	20%	14%	17%	25%	6%	4.25	1.65
(f) We establish standards and targets for performance measures.	1%	6%	15%	25%	22%	25%	6%	4.58	1.38
(g) We continually review and upgrade our SLA / performance measurement system	10%	18%	7%	14%	25%	19%	7%	4.13	1.83
(h) Our staff have the analytical skills needed to effectively analyse the data	1%	7%	10%	17%	28%	22%	15%	4.90	1.50
(i) We have no trouble compiling and distributing the data from our PMS in a timely manner	1%	11%	11%	18%	17%	24%	18%	4.81	1.67
(j) Our performance measures are not ambiguous and confusing	4%	4%	7%	13%	19%	28%	25%	5.22	1.65

Table 5: PMS Effectiveness – Validity

As detailed in table 5 above, answers to the validity questions shows that, around 82% of the respondents indicated their staff had all the analytical skills needed to effectively analyse the data, 85% felt their performance systems were not ambiguous or confusing. Respondents also

reported that their performance measurement systems were easy to understand (72%), distributed in a timely manner (76%) and were based on important data (76%) and targets (78%). When compared to best practice procedures, only 48% of the respondents nominated that they benchmark their performance measures with other similar organisations and only 65% of the sample continually reviewed their service level agreements.

Legitimacy	Never						Usually	Mean	Std. Dev.
	1	2	3	4	5	6	7		
(a) We involve managers in the development / design of our SLA / performance measures	1%	7%	14%	14%	25%	28%	10%	4.79	1.49
(b) We involve the vendors in the development / design of our SLA / performance measures.	1%	7%	7%	16%	20%	40%	9%	5.00	1.45
(c) We involve outside consultants / specialists in the development / design of our SLA / performance measures.	8%	21%	13%	11%	23%	21%	3%	3.93	1.76
(d) We use a third party to verify our SLA/ performance measures.	27%	34%	7%	13%	6%	11%	3%	2.82	1.80
(e) We involve user departments in the development / design of our SLA / performance measures.	13%	11%	13%	14%	21%	24%	4%	4.08	1.82
(f) We do not have trouble getting managers to support our performance measurement system.	3%	8%	14%	30%	8%	24%	13%	4.55	1.63
(g) We do not rely solely on the SLA / performance measures produced by the vendor.	7%	14%	10%	11%	17%	21%	20%	4.59	1.95
(h) We use customer satisfaction surveys to help determine and improve our SLA / performance measurement systems.	8%	13%	7%	15%	15%	17%	24%	4.63	1.99

Table 6: PMS Effectiveness – Legitimacy

When developing the performance measurement system and specifying the service level requirements, it is important that they be based on the actual business needs of the users, rather than the traditional computer metrics that are often used in IT outsourcing arrangements (Streib and Poister, 1999; Warr, 2006). The findings indicate that both vendors and users were involved in the development and design of the performance measurement systems and service level agreements (see table 6 above). Based on the principle of controllability and acceptance of measures (i.e. legitimacy), respondents stated that they involve both vendors (84%) and managers (78%) in the development and design of both service level agreements and performance measures. They also indicated that they used customer satisfaction surveys (72%) and did not rely solely on reports produced by the vendors (69%). Finally, with regard to functionality (see table 7 below), respondents felt their performance measurement systems were useful by enhancing the level of control over the vendor (85%), clearly specifying the rights and responsibilities of all parties (82%) and help motivate the vendor and agency to achieve their results (77%). Respondents did not consider the performance measurement system helped improve the level of accountability (68%) and the level of employee focus on organizational goals (65%).

Functionality	Not at All						Extensively	Mean	Std. Dev
	1	2	3	4	5	6	7		
(a) Our performance measurement system helps improve the relationship between the vendor and the agency.	3%	15%	20%	14%	17%	25%	6%	4.39	1.41
(b) Our performance measurement system enhances the level of communication between the vendor and the agency.	3%	6%	15%	21%	30%	24%	1%	4.55	1.35
(c) Our performance measurement system / service level agreement helps to clearly specify the rights and responsibilities of all parties involved.	1%	7%	10%	17%	28%	22%	15%	4.73	1.37
(d) Our performance measurement system helps improve the quality of decisions or decision-making capacity	8%	6%	14%	11%	36%	24%	1%	4.52	1.47
(e) Our performance system provides improvements in service quality	1%	6%	15%	25%	22%	25%	6%	4.64	1.31
(f) Our performance measurement system helps improve the level of employee focus on organisational goals	10%	18%	7%	14%	25%	19%	7%	3.57	1.45
(g) Our performance measurement system helps improve the level of accountability of individual managers	3%	14%	15%	24%	25%	17%	3%	3.83	1.48
(h) Our performance measurement system enhances the level of control over the vendor.	4%	4%	7%	13%	19%	28%	25%	4.75	1.60
(i) Our performance measurement system motivates the vendor and agency to achieve results and to improve performance.	1%	11%	11%	18%	17%	24%	18%	4.68	1.59

Table 7: PMS Effectiveness – Functionality

Table 8 below shows the combined average for each of the three variables mentioned above. On average, a majority of the respondents indicated that the score for all three variables was above 4.

	N	Minimum	Maximum	Mean	Std. Dev.
Validity	72	1.8	6.4	4.4	1.03
Legitimacy	71	2.2	6.4	4.3	1.08
Functionality	71	2.0	6.2	4.4	1.15

Table 8 Performance Measurement System Effectiveness- Variables

Further exploratory analysis of the PMS Effectiveness construct was conducted using a classification model developed by Bouckaert (2006). This analysis involved classifying the respondent answers provided to each of the twenty seven questions into two levels of effectiveness. Respondent answers to the seven point Likert scale that were equal to or greater than 4 were classified as *high* and answers that were less than 4 were classified as *low*. Based on the results for each of the three dimensions (i.e. validity, legitimacy and functionality), organisations were classified into one of eight groups (Bouckaert, 1993). A system that has high levels in all three dimensions would be considered the ‘optimum system’ and a system that has three low levels would be classified as the ‘worst case’. Between these two extremes there are six other classifications (Bouckaert, 1993). Bouckaert (1993) claims

that an effective performance measurement system should take into account all three dimensions. Results of the analysis and classification of organisations into one of the eight groups is shown in Table 9 below.

	Optimum	Technical and Rhetorical	Technical and Pragmatic	Technical	Rhetorical and Pragmatic	Symbolic	Pragmatic	Worst Case
Validity	H	H	H	H	L	L	L	L
Legitimacy	H	H	L	L	H	H	L	L
Functionality	H	L	H	L	H	L	H	L
Respondent Numbers	33	1	7	5	1	3	8	10
Percentage (%)	49%	1%	11%	7%	1%	4%	12%	15%

Table 9 PMS Effectiveness (Respondents < 4 = Low and > 4 = High)
Based on Bouckaert's (1993) eight group classification model.

As indicated in Table 9 above, 49% of the sample organisations performance measurement systems were classified as 'optimum'. Only 15% of the sample organisations had lows in all three dimensions and was therefore classified under the 'worst case' grouping. These results show that approximately 50% of the sample needs to improve the effectiveness level of their performance measurement system. Organisations that have a low level in functionality are of particular concern because their performance systems are considered to be of no use or benefit to the organisation. An effective performance measurement system is one that provides relevant and useful information that assists managers in the decision making process and helps maintain accountability.

As previously mentioned, the design of an effective performance measurement system in IT outsourcing contracts is a complex process due to the intangibility and heterogeneity characteristics of the IT service provided. In many cases, organisations find it difficult to identify the outcomes and an appropriate set of balanced performance measures that will be used to monitor the vendor's performance (Bouckaert, 1993; General Accounting Office, 1997; O'Shea, 2006). GAO (2001, p. 17) states that one of the most important and difficult tasks in the outsourcing process is writing the specific work statements and service levels for the contract and that performance monitoring of the vendor is the weakest link in the process. See also Cullen et al. (1997) and Humphry (2001).

In summary, the survey results clearly show that Bouckaert's three-dimensional model can be applied to IT outsourcing contractual arrangements and therefore can be used by public sector managers to help develop and design effective performance measurement systems. This model not only highlights the 'worst case' models, but also the 'optimum' models. By classifying the performance measurement systems into one of the eight (8) groups, helps to identify the areas that organisations need to focus. The findings of this research support the propositions outlined (P1, P2 and P3). An effective performance measurement system is influenced by the level of validity, legitimacy and functionality included in the design. Finally, the information gathered in this research contributes to the body of knowledge on contract management and performance measurement system effectiveness and will be of significant importance to the IT and accounting professions and public sector agencies that are considering an outsourcing strategy. The development of a conceptual model that that will assist in the design of effective performance measurement systems will ultimately enhance

the level of accountability, improve communication, assist managers in the decision making process and improve the level of control of the vendor.

Summary, limitations and future research

Deciding whether to outsource is a complex issue and does not suit every organisation (Humphry, 2000). This is demonstrated by the high number of outsourcing arrangements that fail to fully achieve their objectives. It is therefore important that managers responsible for the design and operationalisation of performance measurement systems (PMS) are aware of the factors that influence successful management of the IT outsourcing arrangement. Cullen et al. (2000, p. 230) claim that 'the outcome is determined by how it is managed, before and after the contract is signed'.

A greater focus on accountability and outcome based management has increased the awareness now being placed on performance measurement system (PMS) effectiveness in the public sector (Greco, 1997). The design of a PMS is a complex process due to the intangibility and heterogeneous characteristics of the IT service provided. The appropriate design, implementation and use of the PMS is of critical importance to the success of the outsourcing contract (S. Cullen, Seddon and Willcocks, 2005; Lohman, Fortuin and Wouters, 2003; Streib and Poister, 1999). When designing a PMS, managers should be fully aware of the importance, benefits and characteristics of an effective system. A poorly designed PMS may be of no benefit to the organisation and may also result in dysfunctional behaviours (Neely, 2006; Willcocks, Lacity and Fitzgerald, 1995).

Despite the recognised importance of performance measurement to the overall success of IT outsourcing, a review of the literature highlights that research in this area is very limited. Very little research has focused on the effectiveness or usefulness of performance measurement systems in either the private or public sectors. For example, there is no specific research that relates the effectiveness of performance measurement systems to the success of IT outsourcing contracts. Based on a survey of large public sector organisations and universities throughout Australia, this paper develops and explores a three-dimensional model initially developed by Bouckaert (1993) and later refined by Streib and Poister (1999) to measure the effectiveness of performance measurement systems used in public sector IT outsourcing contracts. This research supports Bouckaert's (1993) proposition that an effective performance measurement system needs to include three essential characteristics - be valid, legitimate and functional. This research contributes to the existing literature on the choice and implementation of PMSs used in public sector IT outsourcing service contracts, and provides some practical guidance to public sector managers on the design of effective PMSs. The results also highlight that an effective PMS is a dynamic process that should be continually improved to help manage the outsourcing relationship. The survey results show that Bouckaert's three-dimensional model can be applied to IT outsourcing contractual arrangements and therefore can be used by public sector managers to help develop and design effective performance measurement systems. This research also shows that an effective performance measurement system can help a business to successfully manage the contract, enhance accountability and provide confidence to the organisation that the outsourcing relationship is being effectively managed and is achieving all of its objectives. Finally, this research demonstrates that a sound understanding of Bouckaert's three-dimensional model will assist public sector managers in the design of effective performance measurement systems that will ultimately enhance the level of accountability, improve communication, assist managers in the decision making process and improve the level of control of the vendor.

The limitations associated with restricting the survey data collection to the public sector and IT outsourcing contracts is acknowledged. However, based on the increasing focus on accountability, level of expenditure and limited research in the area, it was envisaged that significant benefits could be achieved from research in this area. To overcome the potential problem of varying answers associated with multiple contracts, organisations were asked to refer to their largest IT outsourcing contract only. The findings in the survey are therefore limited to only one IT outsourcing contract in each organisation. By focusing on only one contract in the survey minimised the level of confusion and helped to improve the validity and reliability of the data collected. It also enabled this research to focus on the contracts with the greatest level of potential risk and associated benefit to the organisation. The results indicate that more work needs to be done on refining the measurement instrument to improve the quality of the results. It is also recommended that the reliability and validity item questions be improved by reviewing the level of ambiguity and deleting any questions that are redundant.

The limitations detailed above highlight a number of potential opportunities for future research. Information in this area will allow practitioners to better understand the important characteristics or dimensions that contribute to the successful design of an effective performance measurement system. This valuable information will ultimately lead to improved management of outsourcing contracts. Based on the limitations identified, it is recommended that this research be extended to include in-depth one-on-one interviews with public sector managers responsible for the design of the performance measurement systems. It is envisaged that interviews would provide a richer and deeper understanding of the variables that influence PMS effectiveness. It is also recommended that the conceptual model be refined to improve the quality of the results.

References

- Andrews, F. (1984). Construct Validity and Error Components of Survey Measures: A Structural Modelling Approach. *The Public Opinion Quarterly*, 48(2), 409-442.
- Australian National Audit Office (2001). Contract Management: Better Practice Guide. Retrieved 10/6/03, from <http://www.anao.gov.au/publications/betterpracticeguide>
- Barrett, P. (1997). *Accountability for Performance in a More Contractually Oriented Public Sector*. Paper presented at the National Public Accounts Conference, Melbourne.
- Barrett, P. (2001). *Evaluation and Performance Auditing: Sharing the Common Ground*. Paper presented at the Australasian Evaluation Society - International Conference, Canberra, October.
- Bouckaert, G. (1993). Measurement and Meaningful Management. *Public Productivity & Management Review*, 17(1), 31-44.
- Brignall, S., and Modell, S. (2000). An institutional perspective on performance measurement and management in the 'new public sector'. *Management Accounting Research*, 11(3), pp. 281-306.
- Cavalluzzo, K., and Ittner, C. (2004). Implementing performance measurement innovations: evidence from government. *Accounting, Organizations and Society*, 29, p243-267.

- Cullen, S., Seddon, P., and Willcocks, L. (2005). Managing Outsourcing: The Life Cycle Imperative. *MIS Quarterly*, 4(1), 229 - 246.
- Cullen, S. K., Seddon, P., Willcocks, L., University of Oxford., University of Melbourne., & Deloitte Touche Tohmatsu. (2001). *Information technology outsourcing practices in Australia : a joint report prepared by Deloitte Touche Tohmatsu, The University of Oxford and The University of Melbourne*. [Sydney]: Deloitte Touche Tohmatsu.
- Currie, W. (1996). Outsourcing In The Private And Public Sectors: An Unpredictable IT Strategy. *European Journal of Information Systems*(4), pp. 226-236.
- Dean, A. M. (2000). *Issues inherent in measuring and monitoring quality in contracted services*. Caulfield East, Vic.: Faculty of Business and Economics Monash University.
- Dean, A. M., and Kiu, C. (2001). *Performance monitoring and quality outcomes in contracted services : an exploratory study*. Caulfield East, Vic.: Faculty of Business and Economics Monash University.
- Dillman, D. (2000). *Mail and Internet Surveys: The Tailored Design Method* (Second ed.). New York: John Wiley & Sons, Inc.
- Domberger, S., Hensher, D., and Wedde, S. (1993). Competitive Tendering Policies In The Public And Private Sectors. *Australian Journal of Public Administration*, 52(4), pp. 401-411.
- Geiger, D., and Ittner, C. (1996). The influence of funding source and legislative requirements on government cost accounting practices. *Accounting, Organizations and Society*, 21(6), 549.
- General Accounting Office. (1997). *Privatization: Lessons Learned by State and Local Governments. Report to the Chairman, House Republican Task Force on Privatization*.
- Greco, J. (1997). Outsourcing: The New Partnership. *Journal of Business Strategy*(July-August), pp. 48-54.
- Gupta, U., and Gupta, A. (1992). Outsourcing the IS Function: Is it necessary for your organisation? *Information Systems Management*, 9(3), 44-50.
- Hall, C., and Rimmer, S. J. (1994). Performance Monitoring and Public Sector Contracting. *Australian Journal of Public Administration*, 53(4), pp. 453-461.
- Halvey, J. K., and Melby, B. M. (2000). *Business process outsourcing : process, strategies, and contracts*. New York: John Wiley.
- Humphry, R. (2000). *Review of the Whole of Government Information Technology Outsourcing Initiative*. Canberra.
- Industry Commission. (1996). *Competitive Tendering and Contracting by Public Sector Agencies : Overview*: Australian Government Publishing Service, Melbourne.

- Ittner, C., and Larcker, D. (1998). Innovations in performance measurement: Trends and research implications. *Journal of Management Accounting Research*, 10, pp. 205-238.
- Joint Committee of Public Accounts and Audit (JCPAA). (2000). *Contract Management in the Public Service* (No. 379).
- Kaplan, R., and Norton, D. (1997). Why Does Business Need A Balanced Scorecard? *Journal of Cost Management*(May-June), pp. 5-10.
- Kern, T., and Willcocks, L. (2000a). Contracts, Control and 'Presentation' in IT Outsourcing: Research in Thirteen UK Organisations. *Journal of Global Information Management*, 8(4), pp. 15-42.
- Kern, T., and Willcocks, L. (2000b). Exploring information technology outsourcing relationships: theory and practice. *Journal of Strategic Information Systems*, 9, pp. 321-350.
- Lacity, M., and Willcocks, L. (2000). *Inside Information Technology Outsourcing: A State of the Art Report*: Oxford, UK, Templeton College.
- Lacity, M., and Willcocks, L. (2001). *Global Information Technology Outsourcing: In Search of Business Advantage*. Chichester: John Wiley & Sons Ltd.
- Lacity, M., Willcocks, L., and Feeny, D. (1995). IT Sourcing: Maximize Flexibility and Control. *Long Range Planning*, 28(4), 128.
- Langfield-Smith, K., and Smith, D. A. (2004). Management Control Systems and Trust in Outsourcing Relationships. *Management Accounting Research*, 14(3), 281-307.
- Lawton, A., McKevitt, D., and Millar, M. (2000). Coping with Ambiguity: Reconciling External Legitimacy and Organizational Implementation in Performance Measurement. *Public Money & Management*, pp. 13-19.
- Lohman, C., Fortuin, L., and Wouters, M. (2003). Designing a performance measurement system: A case study. *European Journal of Operational Research*, Article in Press.
- MAB/MIAC. (1997). *Before you sign on the dotted line...., ensuring contracts can be managed*. (No. 23): Commonwealth of Australia.
- McFarlan, F. W., and Nolan, R. L. (1995). How to manage an IT outsourcing alliance. *Sloan Management Review*, 36(2), 9.
- Neely, A. (2006). The measurement of performance search for meaningful Effective performance measurement is notoriously difficult to get right, but many organisations view such frameworks as an essential tool for assessing the obligations of individual employees. *Financial Times*, pp 5 - 10.
- Neely, A., and Bourne, M. (2000). Why Measurement Initiatives Fail. *Quality Focus*, 4(4), 3.
- Neely, A., Bourne, M., and Kennerley, M. (2003). Dysfunctional performance through dysfunctional measures. *Cost Management*, 17(5), 41-45.

- Neuman, W. (2000). *Social Research Methods - Qualitative and Quantitative Approaches* (4th Edition ed.): Allyn and Bacon, Massachusetts, USA.
- O'Shea, M. (2006). Outsourcing services - finding the sweet spot. *Contract Management in Practice*, 3(2), pp. 17-20.
- Poister, T., and Streib, G. (1999). Performance Measurement in Municipal Government: Assessing the State of the Practice. *Public Administration Review*, 59(4), pp. 325-335.
- Roberts, S. (1999). In defence of the survey method: An illustration from a study of user information satisfaction. *Accounting and Finance*, 39(1), 53-77.
- Rouse, A. (2002). *Information Technology Outsourcing Revisited: Success Factors and Risks*. Unpublished PhD, The University of Melbourne.
- Simons, R. (2000). *Performance Measurement & Control Systems for Implementing Strategy: Text & Cases*. New Jersey: Prentice-Hall.
- State Supply Commission, W. A. (1997). *Casebook of Good Practice in WA - Effective Contracting for Services*. Perth, Western Australia.
- State Supply Commission, W. A. (2002). The Contract Process Guidelines. Retrieved 20/4/04, from http://www.ssc.wa.gov.au/pol_guide/policies/contract.html
- Streib, G., and Poister, T. (1999). Assessing The Validity, Legitimacy, And Functionality of Performance Measurement Systems in Municipal Governments. *American Review of Public Administration*, 29(2), 107-123.
- Warr, R. (2006). Options for securing IT service provider performance. *Contract Management in Practice*, 3(2), pp. 21 - 24.
- Willcocks, L., Lacity, M., and Fitzgerald, G. (1995). Information technology outsourcing in Europe and the USA: Assessment issues. *International Journal of Information Management*, 15(5), 333.