PUBLIC ADMINISTRATION AND DEVELOPMENT

Public Admin. Dev. **25**, 115–124 (2005)

Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/pad.356

PERFORMANCE MEASURES IN TAX ADMINISTRATION: CHILE AS A CASE STUDY

PABLO SERRA*

Universidad de Chile, Santiago, Chile

SUMMARY

The article proposes a set of tax administration performance measures and contrasts them with measures actually used by the Chilean tax administration agency. The goals assumed for the tax administration agency (TA) are to maximize tax revenue collection and provide quality services to taxpayers. Ideal performance measures (PMs) would measure the deviation of actual outcome from a best-practice standard, given the value of all variables affecting organisation performance that are outside management control. The key challenge is to build and calculate these best-practice outcomes. In Chile the PM in use, for the first goal, is the ratio of actual to potential tax revenue collection. This PM does adjust revenue collection for variations in the tax structure and rate, but it fails to control other variables that affect performance such as the TA budget and per capita income. The PM in use, for the second goal, is taxpayer satisfaction measured through sample surveys. This seems the appropriate PM, as quality of taxpayer services depends directly on the TA efforts to improve them. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS — tax administration; performance measures; Chile

INTRODUCTION

This article offers performance measures for the tax administration agency (TA), and contrasts them with those in use by the Chilean tax administration agency. Although normally the central authority has control over TA, the latter could set its own agenda if not closely monitored by the government.[1] Hence, the need of the central authority to monitor the performance of TA. However, there has been little work directed at this subject, which is probably explained by the difficulty in defining the TA best-practice standards.

Since 1990 Chilean governments have placed increasing emphasis on improving the performance of public bodies. In 1994 the process gathered pace, and within a year 31% of public bodies had defined performance targets, and by 1997 this number had grown to 80%. Moreover, since 1998, public sector employees have been receiving bonuses linked to the achievement of institutional goals. This effort is part of a worldwide trend to foster performance management in public agencies.[2]

Performance management, also labelled as New Public Management (Hood, 1991), is used to describe a management culture that emphasises effectiveness and accountability in public agencies. Their advocates assume that adoption of management tools prevalent in the private sector could reduce wasteful public sector expenditure.[3]

*Correspondence to: P. Serra, República 701, Santiago, Chile. E-mail: pserra@dii.uchile.cl

[1] A semi-independent tax administration institutional design—the Revenue Authority—has been adopted by a number of countries (Devas *et al.*, 2001).

[2] The British public sector reform that began in 1982 (the Financial Management Initiative) was closely followed by New Zealand and Australia. In 1993 the American Government introduced the Performance Results Acts requiring all federal agencies to develop 5-year strategic plans linked to measurable outcomes.

[3] Critics of performance management, however, argue that strict businesslike methods cannot be used in public organisations. They argue that public sector performance is hard to measure as many performances are invisible, unintended or contradictory. Issues are ambiguous, leaving space for interpretation, and in this case the measurement logic does not work (for a summary of these arguments see Noordegraaf and Abma, 2003).

Underlying this development is the concept of performance measurement, where stakeholders use performance measures (PMs) for accountability, evaluation and control,[4] complemented with incentive schemes to encourage efficiency.

Another milestone of the New Public Management is the centrality of the citizen-client. Regarding citizens, public managers must find what their needs are and the measures to evaluate delivery. In addition to the government and the citizens that use the service, public managers face other stakeholders including their employees, Congress and professional organisations. Performance measurement should bear in mind all stakeholders and should be responsive to their objectives and interests.

The process of developing a PM must start from a clear understanding of the organisation's policy goals. The multiple and sometimes vague goals of public institutions, in conjunction with the diversity of stakeholders, makes this an even more crucial stage than in private enterprises.[5] Once the agency's mission and objectives are established, the next step is to develop PMs that measure fulfilment of the corresponding goals. The variety of stakeholders leads to multiple goals that may even be mutually conflicting. In turn, the evaluation of multi-goal agencies requires a set of PMs involving trade-offs (Propper and Wilson, 2003).

PMs could measure inputs, outputs, efficiency (output per unit of input) and outcomes, which need to be analysed in relation to a target, benchmark or some other relevant indicator. Output measures are most useful for agency managers. The standard procedure is to compare latest values against those of previous years, since interagency comparisons pose a number of challenges (Hatry, 1999).[6] However, output measures that ignore quality dimensions provide an incomplete picture of performance. For instance, the knowledge of the number of audits undertaken by the tax administration would be of little use unless their thoroughness is also assessed.[7]

Stakeholders, in turn, care mostly about outcomes; but using raw outcome measures as PMs is inappropriate, since, besides the performance of the agency, they tend to reflect a broad range of external factors over which organisations have little or no control, such as the characteristics of the customers they serve and environment in which they operate. The solution is to use Adjusted Performance Measures (APMs) purged of external factors that affect the organisation's performance. However, these are not straightforward to construct, and once constructed they need to be interpreted carefully (Rubenstein *et al.*, 2003).

Adjusted Performance Measures can be designed as the deviation of the actual outcome from a best-practice outcome, given the value of all variables affecting organisation performance that are outside management control. When cross-section data for a set of comparable agencies or a long historical time series for a single agency are available, techniques such as data envelopment analysis (DEA), introduced by Charnes, Cooper and Rhodes (1978) or the stochastic frontier model, developed by Aigner *et al.* (1977), can be used to estimate the best-practice outcome. These techniques are information-intensive, however, and the necessary information may not be available.

The goals of the Chilean tax administration are to maximise tax revenue, while minimising compliance costs. For the first goal, the Chilean Internal Revenue Service (SII) mostly uses compliance rates as PMs, which adjusts revenue collection for changes in the tax base. Tax compliance, however, has shortcoming as a PM: it does not adjust for changes in per capita income or for changes in the TA budget, factors that affect tax revenue. As an

[4]PMs also are an effective internal management tool. Public managers use PMs to (i) evaluate, (ii) control, (ii) budget, (iv) motivate, (v) promote, (vi) celebrate and (vii) learn—all of which are means to achieve the ultimate goal of improvement (Behn, 2003).

[5]Formulating a clear coherent mission, strategy and objectives is first in Kravchuck and Shack's (1996) list of 10 principles for effective performance management design.

[6]Nevertheless, appraisals of customer satisfaction and measures of productivity in routine processing operations are amenable to interagency analysis.

[7]The 2003 Annual Performance Plan of the US Internal Revenue Service reports 75 PMs. Most of these are output measures and quality indices linked to some of these outputs, the latter based on statistically valid samples of cases handled by each operational unit, which are reviewed and scored by a specialised internal IRS group. The IRS also reports customer satisfaction measures for a wide range of services, and includes a few measures of service standards (e.g. percentage of investigation solved on time), one employee satisfaction measure and one outcome measure (total enforcement revenue).

alternative PM we propose using the difference between actual revenue collected and the maximum feasible revenue collection, given the value of factors outside TA control. In this article we use DEA to compute this difference.

Taxpayers—the other major stakeholders in the tax administration—care about the quality of the services delivered by the TA. The Chilean TA has used an assortment of measures of service quality, but it has not been consistent. During the 1990s it sample-surveyed customers to measure their satisfaction with services, but it has since discontinued the practice. Nowadays it relies more on direct measurements, which provide more objective measures but ignore customer perceptions. The Chilean TA would probably benefit from a more stable and comprehensive set of indicators to measure the quality of their services.

The rest of this article is organised as follows. Section 2 analyses PMs for tax administration; Section 3 assesses the use of PMs by the Chilean tax administration and the final section draws conclusions.

PERFORMANCE MEASURES IN TAX ADMINISTRATION

In what follows, we consider PMs for the TA focusing on agency-wide measures that evaluate the overall performance. We assume that the government sets tax policy, i.e. the tax structure and the TA budget, but delegates responsibility for collecting taxes to the TA. Accordingly, it would seem natural to set maximisation of tax revenue collection as the TA agency-wide goal. The government, however, is not the only principal for the tax administration; its other stakeholders include taxpayers, tax agency employees and tax advisors.

Outcomes achieved by any public agency depend on the behaviour of citizens, and this is particularly true in the case of tax administration. There is empirical evidence that voluntary compliance by taxpayers improves when the TA provides better services (Wallschutzky, 1984). Thus minimising compliance costs—the costs borne by citizens in complying with the tax laws (book-keeping, costs of tax advisors, etc.)—should also be a goal for TA.[8] This is not a minor point, considering that many authors have stressed the fact that people pay more in taxes than suggested by the standard theory of compliance based solely on deterrence. Thus meeting taxpayers' needs is essential for public cooperation and, ultimately, voluntary compliance. The same is true for tax advisors. Hansford and Hasseldine (2002) argue that if the TA and tax advisors develop an adversarial relationship, then current and future non-compliance may increase.

Not surprisingly the U.S. Internal Revenue Service (IRS) states that its mission is to provide taxpayers top quality service by helping them understand and meet their tax responsibilities and by applying the tax law with integrity and fairness to all. And Canada's Customs and Revenue Agency (CCRA) avows that it, 'operates on the fundamental belief that its clients are more likely to comply with the law if they are treated fairly and have the information, advice, and other services they need to meet their obligations'.

Although others goals could be considered, maximisation of tax revenue collection and minimisation of compliance costs seem to be the key ones. Nonetheless, they can be mutually conflicting. If the only goal of TA was to maximise tax revenue collection, it would allocate its budget such that the last dollar spent on enforcement activities produces the same tax revenue as the last dollar spent on improving taxpayer services. But the last dollar spent on improving services will have the additional benefit of also lowering compliance costs. As there is no information of how these two goals trade off against each other, PMs need to be constructed for both of them.

Minimisation of compliance costs

For the goal of minimising compliance costs, the ideal PM would be the deviation of the actual compliance cost from its minimum feasible level; but measuring the latter is an impossible task at the present time. Unadjusted compliance costs do not provide a good measure of the TA performance since they are affected by aspects beyond

[8]Compliance costs are far from negligible. In the U.S. they approach 4.5% of personal individual tax revenues. Guyton *et al.* (2003) estimate that 125.9 million individuals spent an average 25.5 hours and U.S. $149 filling tax returns during year 2000. This year individual tax collection amounted to 1,140 billions of dollars. Assuming that the monetised value of one hour spent for tax compliance was U.S. $10 leads to this figure.

its control such as the complexity of the tax structure and the quality of tax legislation. In contrast, the quality of taxpayer services depends directly on the TA efforts. Taxpayer service quality is usually measured through customer satisfaction surveys. However, results from taxpayers' sample surveys need to be complemented with more objective data, such as percentage compliance with service standards, given that people's perceptions are not necessarily based on their experience.

Canada's CCRA has developed 45 taxpayer service standards covering all major programmes. For instance, the service standard for counter-service waiting time is 20 minutes, the target is 85% served within 20 minutes and in 2001–2002 the 85% standard was met. In the client-resolution programme the service standard is to get back to customers within 15 days, the target is 100% but the standard was only satisfied in 66% of cases. In this process it has heavily relied on customer consultation to identify the most important service elements and customer expectations for service delivery.

Maximisation of tax revenue collection

The main outcome of TA is tax revenue collection. Hence the ideal PM would be the percentage deviation of the actual revenue collection from a maximum achievable outcome, obtained adjusting for all external factors that affect the performance. Tax evasion theory helps in identifying those factors. The traditional theory of compliance asserts that total revenue collection grows with the tax base, the tax rate, per capita income, the size of the penalty for evading taxes and the likelihood of evaders being detected (Allingham and Sandmo, 1972).

The probability of detection, which incidentally is quite difficult to measure, depends in turn on both the budget and the performance of TA. Hence it seems more sensible to make tax revenue collection a function of the tax base, the tax rate, the size of the penalty, the per capita income, the performance and the budget of TA. Moreover, this second representation is more general since TA performance also includes positive encouragement for tax compliance through better taxpayer services, for example.

A sensible approach to measure performance is DEA. Here performance in a specific year would be considered efficient if there is no other year or linear combination of years, which produces more of each output for the given inputs. In the appendix we summarise an alternative method: the stochastic frontier model.

Simpler performance measures

Usually lack of data prevents using the data-intensive measures such as DEA. However, there are a number of PMs for the tax revenue maximisation goal, which adjust for some but not all of the factors that affect tax revenue collection. Some widely used APMs adjust only for the tax administration's budget. The simplest one is the quotient between the total amount collected and TA budget.[9] However, this indicator would show an improvement if tax revenue collection were to rise as a result of a higher tax base even if TA performance had not improved.

The number of taxpayers penalised, or the amount collected from fines are often proposed as PMs,[10] but they might be misleading. When tax enforcement efforts increase, taxpayers are likely to respond by reducing evasion. On the other hand, the proportion of evasion detected can also be expected to increase, so the effect of better TA performance on revenue collected from fines is ambiguous. This shows that ignoring the fact that citizens convert the public agency's output into outcomes, could lead to faulty conclusions. A more suitable PM would be the quotient between fines levied and total estimated tax evasion, which normalises for taxpayers' behaviour. Though this is still an incomplete PM as it fails to consider resources devoted to tax enforcement.

The use of tax penalty collection as the basis for measuring performance has also been criticised on grounds of fairness. The IRS Restructuring Act expressly prohibits the use of tax enforcement results to evaluate employees or set production targets. Examples of 'tax enforcement results' that may not be used to evaluate an employee include

[9]The Inter-American Tax Administrators Center estimates the quotient between revenue collection and the tax administration budget for countries in the region. The IRS claims that it is one of the world's most efficient tax administrators since it cost taxpayers 45 cents for each $ 100 of tax revenue collected in 2002.

[10]Cremer *et al.* 1990, suggest maximising the number or amount of fines for a given audit budget as the TA goal. Hunter and Nelson 1996, assume the output of TA to be tax enforcement, and that this can be adequately measured by the additional taxes and penalties assessed by the TA.

the amount assessed or collected. Quantitative goals—even if only applied at the management level—may result in examinations that are unduly focused on raising revenue, rather than verification of tax return data.

A more appropriate APM is the compliance ratio: the ratio of actual to potential tax receipts. For instance, Lewis (2003) measures the administrative performance of property tax revenue collection in Indonesia through this ratio.[11] Potential tax receipts do take account of variations in the tax base and rate, but do not control other variables that affect the actual revenue collection such as changes in per-capita income and the budget of the TA.

CHILE AS A CASE STUDY

The Chilean Internal Revenue Service (SII) is an autonomous decentralised public agency in charge of formulating and executing the tax enforcement strategy. The government therefore has to set performance targets and devise suitable incentives for the SII to respond to them. The Government uses tax revenue collection to rate the performance of the SII. The benchmark is set equal to the revenue collected during the previous year, adjusted by that year's rate of GDP growth times 1.1 (the estimated elasticity of tax revenue collection with respect to GDP). SII employees earn bonuses that depend on fulfilment of this target.

This PM reflects the government's own needs. First, the desire to link employee rewards to uncontroversial indicators that are easy to build from readily available data sources (Hatry, 1999; Behn 2003). Moreover, governments care most about tax revenue collection, since they need this to finance public expenditure; and to decide on the SII budget, an important piece of information is the revenue that an extra peso spent on tax administration would yield.

Though quite useful for its own purposes, the PM set by the government does not provide a comprehensive evaluation of the SII. Although the government's PM accounts for fluctuations in the GDP, the tax base may change also with variations on the tax structure and in the composition of the GDP. It also relates only to one of the SII objectives, namely maximising revenue collection. Accordingly, the SII has developed its own set of PMs to measure organisation's effectiveness based on the institution's objectives.

The stated mission of the SII is to administer the internal taxation system, monitor taxpayers to ensure they comply with the tax laws, and facilitate compliance. Its specific objectives are to improve the agency's efficiency; promote professional and personal development among its staff; reduce levels of tax evasion and avoidance; facilitate tax compliance and improve taxpayer services; ensure equity and legal precision in the application of tax laws; strengthen and develop greater technological capacity to fulfil the mission. The two basic objectives, however, seem to be maximising tax revenue and minimising compliance costs; while the other objectives serve as means for achieving them.

Minimising compliance costs

Although no attempt has been made to measure compliance costs in Chile, customer satisfaction has been assessed since the early 1990s. The SII has mainly relied on taxpayer surveys conducted by external opinion research firms, although more recently it has turned to objective data measures. Image surveys were conducted on a regular basis in the past decade (1990, 1992, 1996 and 2000), and the results obtained from the three most recent are summarised in Table 1. Interviewees were also asked to rate the SII relative to other public and private services. These surveys measure taxpayers' satisfaction, rather than appraising service quality directly.

The SII also commissioned sample surveys among individuals summoned to the SII offices to clarify inconsistencies in their income tax returns in 1994, 1996 and in 1997. Taxpayers were asked to rate different features of SII services, some of which are listed in Table 2. The survey also included one service standard, namely the percentage of taxpayers that completed their procedures in one visit. Similar sample surveys were carried out in 1996 and in 1997 of taxpayers visiting SII offices to undertake a 'life-cycle' procedure (i.e. start up a business, obtain a tax identification number or get invoices stamped). These surveys included one service standard measurement, namely

[11]He decomposes this ratio into the coverage ratio, the valuation ratio and the collection ratio, providing more disaggregate information on the performance of the administration.

Table 1. SII evaluation by taxpayers (% of respondents that grades the SII 5 or above in a scale 1 to 7)

	1992	1996	2000
Modernisation	75.0	60.9	80.7
Agility	52.6	31.4	47.9
Technological level	83.2	62.8	83.0
Professionalism	75.8	58.7	64.0
Efficiency	72.4	51.5	63.6
Honesty	78.9	73.3	81.8
Strictness	83.6	87.6	83.4
Customer servicing	59.0	51.2	56.0
Accuracy	79.1	77.7	74.6
Progress in computerisation	88.2	68.9	89.5

Source: Adimark surveys (SII website).

Table 2. Customer satisfaction with SII workplace services (% of visitors satisfied or highly satisfied)

	Income-tax procedures			Life-cycle procedures	
	1994	1996	1997	1996	1997
Employees	72.4	86.8	83.2	77.9	84.6
Published tax guidelines	70.8	78.2	77.8	74.4	77.5
Overall satisfaction	60.3	58.5	64.1	66.4	72.5
Office quality	45.7	58.5	61.9	78.1	83.9
Counter-line waiting time	31.3	39.6	51.9	48.4	61.5
Equality in treatment	34.0	75.0	74.9	64.8	72.3
Taxpayers that completed the procedure in one visit (%)	68.9	70.1	69.7	63.6	64.3
Taxpayers that spent more than 30 minutes in the procedure (%)				36.6	27.4

Source: Adimark surveys (SII website).

counter-line waiting time. Surveys were later discontinued, leaving the SII without knowledge of customer satisfaction. The SII is currently relying on direct measurements of service standards.

The 'Maximum Waiting Time' project introduced in 1997 guarantees taxpayers a maximum waiting time of 30 minutes for all 'life-cycle' procedures. If this time limit is exceeded, the SII has to take the documents to the taxpayer's address. When a taxpayer arrives at the SII office, he/she takes a ticket from an electronic dispenser, and a counter registers this when he/she starts to be attended. The software used records on-line variables such as the number of staff helping the public, the number of taxpayers attended by each staff per unit of time, and time required to serve taxpayers. Compliance with the 30-minute standard has declined slightly since 2001, but the SII points out that it has attended a larger number of taxpayers on time with the same number of employees. The results are shown in Table 3.

Since the late 1990s SII has made efforts to reduce compliance costs further by using information technology. Tax forms can now be downloaded from the SII website, which also allows taxpayers to carry out most tax transactions, including completing tax returns, paying taxes, receiving information on the status of their income tax returns and making complaints. The SII itself prepares a draft of income-tax return for most taxpayers that can be accessed from the website. In recent years SII has consulted their customers on how to improve Internet services, but the results do not have the formality of a survey.

An indirect measure of the quality of Internet services would be the extent to which they are used by taxpayers. In 1999, 5.2% of all taxpayers filed their income tax returns through the SII website, and this figure had risen to 55% by 2002. In the case of Form 29, which is filed monthly (VAT returns and provisional income-tax payments), the percentage of Internet use is much lower—though also growing fast as shown in Table 4.

Table 3. Counter-line waiting time statistics

	Number of taxpayers assisted (thousands)	%waiting longer than 30 minutes
1998		7.2
1999		0.2
2000	1,439	0.2
2001	1,542	0.6
2002	1,657	1.0

Source: SII, Subdirección de Fiscalización.

Table 4. Use of Internet in tax returns (thousands of returns)

	Form 29			Income Tax Returns		
	2000	2001	2002	2000	2001	2002
Internet	557	602	1,237	467	789	1133
Paper	11,635	11,163	11,175	1365	1107	936
Total	12,192	11,765	12,412	1832	1896	2069

Source: SII, Subdirección de Fiscalización.

Tax revenue collection

The SII uses the compliance rate as a performance measure for the revenue maximisation goal. Measurement is restricted to the VAT. Focusing on VAT is justified on several grounds: first, it is easier to compute the compliance rate; second, it accounts for about 45% of all tax revenue collected; third, VAT evasion reduces the tax base in other tax categories, with about 75% of company income tax evasion explained by evasion of VAT. As personal and company income taxes are integrated, VAT evasion also erodes the personal income tax base.

The SII estimates compliance by calculating potential revenue based on the National Accounts for each year and comparing this with what is actually collected. Broadly speaking, the VAT base is calculated by adding together the different transactions that are subject to VAT. The main component is final consumption, both public and private, of goods and services that are subject to the tax. Apart from this, there is the purchase of VAT-able intermediate and investment goods, which are then used in the production of goods and services that are exempted.

A major advantage in using the compliance rate as a PM is that it provides information about aspects that citizens understand and care about. As Behn (2003) points out, it is important for managers of public bodies to convince citizens that their agency is effective and efficient, for which purpose they need to develop easy-to-understand measures of aspects of performance that many citizens personally care about. The tax evasion rate is also amenable to inter-country comparisons and some of these have been performed.[12] Moreover, some authors maintain that taxpayers' behaviour is influenced by the behaviour of the rest of society (Cullis and Lewis, 1997); in other words they seem more prone to evade taxes in societies that already have a high tax evasion rate. Hence the existing tax evasion rate is information that taxpayers' take into consideration when making their own decisions.

A difficulty with this PM concerns the availability of timely information, since there is a delay in publishing the national account figures that are used in calculating the compliance rate. It takes over a year even to obtain provisional figures, and these are later revised and sometimes altered substantially. Clearly, employee incentives cannot be linked to performance indicators such as these. The major problem with the compliance rate, however, is that it

[12]International comparisons have their limitations, however; for example, the possibility of catching tax evaders depends on various factors including the complexity of the tax system and the information that the TA is authorised to request from taxpayers and third parties. Nonetheless, significant differences in performance would need to be explained, and the explanation could lead to a change—for instance, simplification of an overly complicated tax system.

Table 5. Performance in revenue collection

Year	Tax rate	SII budget	Potential VAT tax revenue	Actual VAT tax collection	Compliance rate %	DEA performance measure
		Current Ch$				
1988	18.00	4,440	518,839	375,828	72.4	88.4
1989	16.00	4,894	619,954	452,750	73.0	92.2
1990	17.00	6,028	834,349	602,291	72.2	96.3
1991	18.00	8,626	1,181,926	875,605	74.1	97.8
1992	18.00	11,466	1,554,605	1,190,190	76.6	100.0
1993	18.00	15,608	1,849,531	1,515,285	81.9	100.0
1994	18.00	18,778	2,197,500	1,695,860	77.2	94.1
1995	18.00	22,013	2,612,959	1,990,068	76.2	93.0
1996	18.00	25,763	2,995,280	2,328,578	77.7	94.8
1997	18.00	26,979	3,248,164	2,619,917	80.7	99.3
1998	18.00	33,465	3,560,788	2,710,816	76.1	92.4
1999	18.00	39,641	3,580,041	2,793,349	78.0	93.8
2000	18.00	47,423	3,914,990	3,086,629	78.8	94.3
2001	18.00	53,200	4,245,324	3,404,920	80.2	95.7
2002	18.00	63,620	4,443,538	3,700,460	83.3	98.8
2003	18.25	64,821	4,788,331	4,036,873	84.3	100.0

Sources: SII and own calculations (last column).

does not adjust for all factors affecting tax revenue collection, since it ignores per capita income and tax administration budget.

We use data envelopment analysis (DEA) to estimate the maximum feasible tax revenue collection, and then compare the resulting number with actual collection. The inputs that go into tax revenue collection are the number of employees and the potential tax revenue. Results are shown in Table 5. The correlation between both PMs—the compliance rate and DEA—is a moderate 0.63. The compliance rate rises sharply in the early 1990s, but the DEA performance measure does not. Chile's strong improvement in terms of VAT compliance is partly attributable to the SII budget's strong growth. Hence using the compliance rate as a PM seems to overestimate the effectiveness gains achieved by the SII during the 1990s.

The Chilean experience also provides a vivid example of the negative effects of a poorly designed PM. Prior to 1991, the performance of the SII regional offices was rated by the level of their penalty assessments. This encouraged tax inspectors to make frivolous estimates, assessing penalties that did not exist or imposing overvalued fines that subsequently had to be downgraded. From 1992 onwards, performance has been measured through net penalty appraisals, i.e. final assessments after elimination or reduction by SII regional director. Penalty assessments by auditors fell dramatically in 1992, whereas fines collected increased, as shown in Table 6. The last row of the table reports an adjusted outcome measure, namely the quotient between fines collected and total estimated

Table 6. Chile: VAT enforcement performance (Millions of Chilean pesos, constant January 1999 prices)

	1991	1992	1993	1994	1995	1996
VAT sanction collection	7,640	11,911	10,680	13,530	12,368	13,806
VAT auditors' assessment	55,147	40,126	36,746	42,834	40,387	40,043
VAT penalty collection (%)	13.9	29.7	29.1	31.6	30.6	34.5
VAT non-compliance amount	636,360	610,531	515,143	577,182	646,982	673,272
Fines levied as a % of VAT non-compliance amount	1.20	1.95	2.07	2.34	1.91	2.05

Source: SII, Subdirección de Estudios.
Note: As from 1997, the SII changed its methodology for measuring revenue collected as a result of inspection. The new methodology involves measurement by plans and programs rather than by taxes, as had been the case up to 1996.

evasion. This indicator posted a strong gain in 1992, which has been followed by mild fluctuations since 1993. This example illustrates the potential PMs have to distort organisational behaviour, as individuals respond to them in ways to maximise their own utility or benefit and not necessarily always as expected by the PM designers.

CONCLUSIONS

This article draws on the literature and experience in Chile to offer lessons for policymakers wishing to implement performance measurement in the tax administration agency. The first step in this process should be to develop a clear understanding of the goals of the organisation, and then formulate outcome measures associated with these goals. The next step should be to define best-practice outputs. Theory and/or expert knowledge should serve to identify the external factors that affect the organisation's performance, and statistical or optimisation methods are proposed to control outcomes for those factors. Stakeholder diversity leads to multiple goals that may be mutually conflicting to some extent. In our view, policymakers and not public managers should be responsible for making trade-offs between the PMs associated with these goals, ideally through dialogue with the other stakeholders.

Lack of quantitative data (or knowledge) may tempt policymakers to employ simplified PMs, but this practice can result in spurious conclusions. For instance, the compliance ratio is frequently used as a tax administration PM; but this criterion ignores the fact that other variables, such as changes in per capita income and the budget allocated to TA, also affect compliance. One solution would be to use the compliance rate as a PM for the tax administration, but at the same time scrutinise the behaviour of other variables affecting compliance and make a qualitative assessment of performance considering those aspects.

As many authors have pointed out, simplified PMs also have the potential to distort organisational behaviour, since agencies will tend to maximise their PM score instead of fulfilling institutional objectives. In Chile, evaluation of tax auditors based on the level of fines assessed gave rise to many frivolous assessments causing significant costs to taxpayers during the early 90s. The implementation of simplified PMs thus needs to be preceded by a thorough analysis of their implications for the behaviour of the officials concerned. One way of mitigating this problem is to include the satisfaction of all relevant stakeholders among the performance measures.

Some public agencies are more amenable to performance measurement than others.[13] As tax administration is probably at one end of the spectrum, it would be misleading to use it as a basis for quantifying the effort needed to implement performance measurement in the public sector generally. In some agencies research is needed to define the goals and the corresponding APMs. Implementing performance measurement could therefore be a resource- and time-consuming process, but this should motivate an earlier start to implementation rather than delay, assuming that the resources needed to perform the task, both financial and otherwise, are available.

To sum up, many authors argue that performance measurement is essential for both accountability and promoting a culture of effectiveness in the public sector; we agree. Nonetheless, the complexity of measuring public sector performance may lead to simplified PMs being used, which in turn may give rise to faulty conclusions or produce distortions in the behaviour of officials. Identifying all the factors affecting the agency's relevant outputs can reduce these potential risks, but this in turn may require wide-ranging research to build a theory in situations where none exists.

REFERENCES

Aigner D, Knox Lovell C, Schmidt P. 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* **6**(1): 21–37.

Allingham M, Sandmo A. 1972. Income tax evasion: a theoretical analysis. *Journal of Public Economics* **1**(3–4): 323–338.

[13]Noordegraaf and Abma (2003) define canonical practices as those where issues are known and standards relatively uncontested. In such cases performance can be measured and repetitive performances can be compared. At the other extreme non-canonical practices exist where processes are non-routine, fuzzy, innovative and conflicting. Knowledge of issues can be assimilated to identifying factors that affect performance, while uncontested standards would be akin to agreeing on APMs that control for those factors.

Behn R. 2003. Why measure performance? Different purposes require different measures. *Public Administration Review* **63**(5): 586–606.
Charnes A, Cooper W, Rhodes E. 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research* **2**(6): 429–444.
Cremer H, Marchand M, Pestieau P. 1990. Evading, auditing and taxing: the equity-compliance trade-off. *Journal of Public Economics* **43**(1): 67–92.
Cullis J, Lewis A. 1997. Why people pay taxes: from a conventional economic model to a model of social convention. *Journal of Economic Psychology* **18**(2–3): 305–321.
Devas N, Delay S, Hubbard M. 2001. Revenue authorities: are they the right vehicle for improved tax administration? *Public Administration and Development* **21**(3): 211–222.
Guyton J, O'Hare J, Stavrianos M, Toder E. 2003. Estimating the compliance cost of the U.S. individual income tax. *National Tax Journal* **56**(2): 673–688.
Hansford A, Hasseldine J. 2002. Best practice in tax administration. *Public Money and Management* **22**(1): 5–6.
Hatry H. 1999. Mini-symposium on intergovernmental comparative performance data. *Public Administration Review* **59**(2): 101–104.
Hood C. 1991. A public management for all seasons? *Public Administration* **69**(1): 3–19.
Hunter W, Nelson M. 1996. An IRS production function. *National Tax Journal* **49**(1): 105–115.
Kravchuck R, Shack R. 1996. Designing effective performance-measurement systems under the government performance and result acts of 1993. *Public Administration Review* **56**(4): 348–358.
Lewis B. 2003. Property tax in Indonesia: measuring and explaining administrative (under) performance. *Public Administration and Development* **23**(3): 227–239.
Noordegraaf M, Abma T. 2003. Management by measurement? Public management practices amidst ambiguity. *Public Administration* **81**(4): 853–871.
Propper C, Wilson D. 2003. The use and usefulness of performance measures in the public sector. *Oxford Review of Economic Policy* **19**(2): 250–267.
Rubenstein R, Schwartz AE, Stiefel L. 2003. Better than raw: a guide to measuring organizational performance with adjusted performance measures. *Public Administration Review* **63**(5): 607–615.
Wallschutzky I. 1984. Possibly causes of tax evasion. *Journal of Economic Psychology* **5**(4): 371–384.

APPENDIX

We have made tax revenue collection, R, a function of the tax base, B, the per capita income, y, the tax rate, t, the size of the penalty for evading taxes, s, the budget of the tax administration, g, and the performance of the TA, α:

$$\mathrm{R} = \mathrm{g}(\mathrm{B}, \mathrm{y}, \mathrm{g}, \mathrm{t}, \mathrm{s}, \alpha) \tag{1}$$

When a long time series is available, the aggregate revenue collection function can be estimated resorting to econometric methods. Tax revenue collection is regressed on those factors affecting performance that TA does not control. Assuming a log-log specification for Equation (1) we write:

$$\log \mathrm{R} = \beta_0 + \beta_{\mathrm{B}} \log \mathrm{B} + \beta_{\mathrm{y}} \log \mathrm{y} + \beta_{\mathrm{t}} \log \mathrm{t} + \beta_{\mathrm{s}} \log \mathrm{s} + \beta_{\mathrm{g}} \log \mathrm{g} + \beta_{\alpha} \log \alpha + \mathrm{v} \tag{2}$$

where v denotes the error term. Omitting variable α from Equation (2) leads to:

$$\log \mathrm{R} = \beta_0 + \beta_{\mathrm{B}} \log \mathrm{B} + \beta_{\mathrm{y}} \log \mathrm{y} + \beta_{\mathrm{t}} \log \mathrm{t} + \beta_{\mathrm{s}} \log \mathrm{s} + \beta_{\mathrm{g}} \log \mathrm{g} + \mathrm{v} \tag{3}$$

In this context, Rubinstein *et al.*'s (2003) proposition would be to regress Equation (3) and use the residuals as an APM. The rationale for this is that as the effectiveness variable α has been omitted from Equation (3), the residuals encompass the effects of this variable. Assuming α is not correlated with the other explanatory variables, which seems reasonable, its omission does not bias the estimation, but it could result in an inefficient estimate of performance since the residuals are unable to disentangle the effect of α on tax revenue collection from that of other components of the error term.

The stochastic frontier production estimation method, developed by Aigner *et al.* (1977), specifies the error term as being made up of two components, one normal and the other from a one-sided distribution, where the one-sided error term represents technical inefficiency. The corresponding equation for estimation is:

$$\log \mathrm{R} = \beta_0 + \beta_{\mathrm{B}} \log \mathrm{B} + \beta_{\mathrm{y}} \log \mathrm{y} + \beta_{\mathrm{t}} \log \mathrm{t} + \beta_{\mathrm{s}} \log \mathrm{s} + \beta_{\mathrm{g}} \log \mathrm{g} + \mathrm{v} - \mathrm{u} \tag{4}$$

The data are expressed in log form, so u measures the percentage deviation of each observation from the frontier and 100-v becomes a performance measure.

