



香港城市大學
City University of Hong Kong

專業 創新 胸懷全球
Professional · Creative
For The World

CityU Scholars

Analyzing the Complexity of Performance Information Use Experiments with Stakeholders to Disaggregate Dimensions of Performance, Data Sources, and Data Types

Walker, Richard M.; Jin Lee, M.; James, Oliver; Ho, Samuel M. Y.

Published in:
Public Administration Review

Published: 01/11/2018

Document Version:
Post-print, also known as Accepted Author Manuscript, Peer-reviewed or Author Final version

Publication record in CityU Scholars:
[Go to record](#)

Published version (DOI):
[10.1111/puar.12920](https://doi.org/10.1111/puar.12920)

Publication details:
Walker, R. M., Jin Lee, M., James, O., & Ho, S. M. Y. (2018). Analyzing the Complexity of Performance Information Use: Experiments with Stakeholders to Disaggregate Dimensions of Performance, Data Sources, and Data Types. *Public Administration Review*, 78(6), 852-863. <https://doi.org/10.1111/puar.12920>

Citing this paper

Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights

Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission

Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy

Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.

**Analyzing the complexity of performance information use: Experiments
with stakeholders to disaggregate dimensions of performance, data sources
and data types**

Richard M. Walker
Laboratory for Public Management and Policy
Department of Public Policy
City University of Hong Kong
Tat Chee Avenue
Kowloon, Hong Kong
rmwalker@cityu.edu.hk

Richard M. Walker is Chair Professor of Public Management at City University of Hong Kong where he is Director of the Laboratory for Public Management and Policy, and Associate Dean (Research and Postgraduate Studies) in the College of Liberal Arts and Social Sciences. Richard is Vice-President of the Public Management Research Association.

M. Jin Lee
Laboratory for Public Management and Policy
Department of Public Policy
City University of Hong Kong
m.jinlee@cityu.edu.hk

M. Jin Lee is Research Fellow of Public Policy at City University of Hong Kong. She is a core member of Laboratory for Public Management and Policy, and manages the experimental methods lab. Her research focuses on public management, decision-making, motivation and policy networking.

Oliver James
Department of Political Science
University of Exeter
o.james@exeter.ac.uk

Oliver James is a Professor of Political Science in the Department of Politics at the University of Exeter. Oliver works on the politics and accountability of public services, citizen-provider relationships, public sector organisation and reform, executive politics and regulation of publicly owned and/or funded bodies and services. He uses a range of methods including survey and field experiments.

Samuel Ho
Psychology Laboratory
Department of Applied Social Sciences
City University of Hong Kong
Hong Kong
munyinho@cityu.edu.hk

Prof. Samuel M.Y. Ho is the Associate Provost (Institutional Initiatives) and a Professor of Psychology at the City University of Hong Kong. Currently, Prof Ho is conducting a series of experiments to examine the effect of different information processing styles on mental health. He is one of the representative figures of positive psychology in Asia, and the Executive Council member of the Clinical Division of the International Positive Psychology Association

Acknowledgements: We acknowledge the support given by co-investigators Peter Hills, Ka Ho Mok, Yetta Wong and Alfred Wu in this research study.

Funding: The research reported in this article was funded by the Public Policy Research Funding Scheme from the Central Policy Unit of the Government of the Hong Kong Special Administrative Region (Project Number: 2015.A1.031.16A) and the National Research Foundation of Korea Grant, the Korean Government (NRF-2011-330-B00194).

**Analyzing the complexity of performance information use: Experiments
with stakeholders to disaggregate dimensions of performance, data sources
and data types**

Abstract

We address important questions about the complex construct of underlying performance information use (PIU) – public service performance. A between-subjects experimental vignette methodology was implemented to answer questions about the effects of emphasizing different dimensions of performance, and the sources and types of performance information among internal and external stakeholders in two service arenas (secondary education and solid waste management) in Hong Kong. The findings indicate common attitudes and agreement across stakeholder groups and services on the merits of archival and external data types. Other results vary by service and between stakeholder groups. The effects of information about effectiveness can depend on its combination with information about efficiency or equity. This complexity needs to be considered when designing information communication to different stakeholder groups.

Practitioner points

- Public service organizations should consider providing targeted and appropriate performance information (PI) that suits the needs of stakeholder groups.
- The effect of providing information about effectiveness depends on its combination with information about efficiency and equity. Providing information about efficiency as well as effectiveness reduced perceived performance in the case of secondary education.
- Whenever possible public service organizations should communicate external and archival PI to stakeholder groups.

Performance information use (PIU) is the most widely diffused management practice to emerge from the New Public Management reform movement. Although the effects and merits of PIU remain contested (Heinrich and Lynn 2000; Pollitt 2006; Van Dooren and Van de Walle 2008) governments themselves have used it to become more performance orientated. The OECD (2005) describes how governments use PIU to enhance planning, reporting and controls systems to improve decision-making. Behn (2003) identifies eight purposes of performance measurement: evaluate, control, budget, motivate, promote, celebrate, learn, and improve. These practices are supported by burgeoning research evidence (George et al. 2016; Moynihan 2008). One aspect of this expanding knowledge base is the growing interest in the interface between the government and citizens (Baekgaard 2015; Baekgaard and Serritzlew 2016; James and John 2007). Government-citizen PIU studies, the focus of our enquiry, have shown how citizens use performance information (PI) to punish or reward politicians in elections (James and John, 2007), that PI matters to citizens when linked to cost information (Baekgaard 2015), that relative PI data affect perceptions of performance (James and Moseley 2014), and that citizens' interpretation of PI are biased by their ideological beliefs (Baekgaard and Serritzlew 2016).

This article contributes towards the knowledge of PIU and is partly motivated by James' (2011, 399) call to assess information "in different service contexts coming from different information sources" and James and Moseley's (2014, 14) recommendation to focus on "multi-dimensional information (for example about inputs, processes, or outputs and different aspects of a service)." Performance is a complex concept and can be operationalized in many ways, reflecting the multiple goals of public organizations. We seek to unpack the concept, thereby making a number of contributions to the study of PIU. The first contribution we make is examining the preferences for different dimensions of performance in different stakeholder groups. Many dimensions of performance include outputs and outcome

measures, of which the most frequently seen in the literature are efficiency, effectiveness and equity (Boyne 2002; Walker, Boyne and Brewer 2010). To date preferences for dimensions of performance have only been examined when investigating the trade-offs of single stakeholder groups. For example, Mikkelsen (2017) found that managers did not make trade-offs between production and process performance when delivering services. Andrews and Van de Walle (2013) found that citizens associated different management practices with different dimensions of performance. However, Grosso, Charbonneau and Van Ryzin (2017) examined how citizens treated outputs and outcomes. They found that although citizens were expected to be especially interested in information about outcomes to hold government accountable for the consequences of their activities, they responded more to output data. We examine stakeholders inside and outside of the public service production function.

The complexity of PI does not end with identifying the different dimensions of performance. Data on each dimension can be communicated from sources internal or external to an organization, and derived from perceptual or archival sources. These alternative operationalizations are important for further disaggregations of the complex performance construct. Some studies have focused on presenting absolute and relative PI to citizens and , how sources of information affect its credibility to citizens (Boyne 2010; Holzer and Kloby 2004; James and Moseley 2014; James and Van Ryzin 2017). However, no studies have examined information sources and types; this forms our second contribution.

Third, we make a novel contribution to the experimental work on this topic by using subject pools beyond those that have dominated the related research which has drawn almost exclusively on U.S. and European populations, especially in the U.K. and Denmark. An overreliance on narrow subject pools has been a source of criticism in, for example, psychology (Arnett 2008; Machery 2010). We seek to correct this imbalance by conducting a series of experiments in Hong Kong. We also build on previous studies by examining two

service areas: education (secondary schooling) and solid waste management (recycling).

The following section reviews the concept of organizational performance, breaking it down into the components we address in a set of experimental studies based upon our research expectations. We follow this with a discussion of the methods. Experiments are advantageous for studying performance measures compared with observational methods. The use of PI is often associated with factors such as satisfaction with services that also affect dependent variables of interest (e.g., perceptions of performance) in a way that conventional statistical techniques, notably multivariate regression, find challenging to separate out due to the difficulty of incorporating adequate control variables (James, Jilke and Van Ryzin 2017). We adopt this methodological approach and implement an Experimental Vignette Methodology (EVM; Aguinis and Bradley 2014). Two experiments were conducted, the first on the dimensions of performance, and the second on sources and types of information. The findings of these studies largely indicate agreement across stakeholder groups and service attitudes on the merits of relative and external data types. However, the evidence points to more complex performance dimensions, with the results varying by service and between stakeholder groups. The discussion develops the implications of these findings for the study of PIU and proposes an agenda for future research.

Performance Information Use: Organizational Performance and Research Expectations

Formal systems that provide PI to internal actors who deliver public services (politicians and managers) and to external stakeholders (citizens and service users) have flourished for many decades in public organizations. Theories on PIU drawn from the economics, political science and public administration fields makes similar arguments on PIU from the perspectives of internal and external actors (Boyne et al. 2003; Moynihan 2008; Sharpe 1970). PI provides citizens and service users with information that allows them to become

more knowledgeable, voice their concerns or exercise choice of service where this is available. This use of information, coupled with responses by politicians and public managers, can help improve the quality of services and make them more responsive to citizens and users (James and John 2007; James and Moseley 2014). Managerially, PI can be seen as part of a cybernetic system that provides information to managers on service strengths and weaknesses and allows action to be taken (Moynihan 2008). Performance systems entail service standards and performance indicators. Targets are placed against these indicators and actions are taken on resulting information about how actual performance compares with targeted levels. These actions can be taken by managers within an organization or by citizens and users, actors external to the organization pressuring managers, and politicians to make the organization respond to their concerns, such as by voicing concerns or by exiting to rival service providers and thereby exerting “market pressures” (James 2011; James and Moseley 2014). Although evidence has pointed to the validity of these systems and regimes as engines of change and improvement (Boyne 2010; Moynihan 2008), research and practice have rarely considered the range of PI that can be made available on service efficiency, effectiveness and equity, or whether the data are standalone or comparative, from internal or external sources, or archival or perceptual. We unpack the concept of organizational performance before outlining the expected relationships.

Organizational Performance

The majority of PIU studies have focused on measures of organizational performance as either dependent variables, for example to assess the effects of managerial actions or organizational structures, or independent variables with information affecting managers or citizens’ perceptions, attitudes, or behavior (Moynihan 2008; James and John 2007; James and Moseley 2014). However, there is considerable complexity in performance measures, and the concept is somewhat controversial in nature (Heinrich and Lynn 2000; Walker,

Boyne, and Brewer 2010). In this section, we review the factors that cause this complexity: the range of conceptual dimensions of performance, the number of interested stakeholders, and the different types and sources of data (Andersen, Boesen and Pedersen 2016; Walker and Andrews 2015; Walker, Boyne and Brewer 2010). Figure 1 is an illustration of the performance construct. The horizontal axis shows stakeholders who produce information (internal or external to the organization) and the vertical axis shows types of data. The cells provide the performance dimensions using the example of effectiveness measured by student attainment in secondary school education. Following data production, both internal and external stakeholders use the information, such as in internal management or external information, to enable user choice, exit from services or communicate with politicians or managers.

[Figure 1 about here]

Performance is a multidimensional construct. Organizations produce a range of outputs and outcomes and a number of models have been proposed (Boyne 2002; Selden and Sowa 2004; Walker, Boyne, and Brewer 2010). The 3Es model includes economy, efficiency, and effectiveness, whereas the input-output-outcome model examines the relationships between inputs, outputs and outcomes. Key outputs include speed, quality and quantity, and key outcomes focus on efficiency, effectiveness, equity, and responsiveness (Boyne 2002). Walker, Boyne, and Brewer (2010) extended this logic to include broader concerns about the governance of public organizations, arguing that several additional dimensions must be included, such as accountability, probity, and respect for human rights. To fully capture the performance achievements of a program or an organization, multiple dimensions of PI should occupy the cells in Figure 1. In practice, data availability often restricts empirical studies and performance reports by public agencies to a limited number of performance indicators. For example, integration studies of management and performance find extensive use of

effectiveness and limited examples of efficiency and equity (Boyne 2002; Walker and Andrews 2015).

An often-cited characteristic of public organizations is their complex external environments, which can influence performance achievement, and stakeholders are one key influence (Pettigrew, Ferlie, and McKee 1992; Thompson 1967). In addition to the managers and employees of an organization, its various stakeholders can include overhead political authorities: other public, private, or non-profit organizations: service users, and citizens, all of whom constitute important components of the external environment. These internal or external stakeholders are shown in Figure 1. Stakeholders can exert power over public agencies in different ways by granting them legitimacy or conferring upon them an urgency to act. They are therefore of interest to the study of organizational achievements (Mitchell, Agle, and Wood 1997). Different stakeholders are likely to express interest in different parts of the “public sector services production function” and prioritize alternative PI (also see Moynihan’s [2008] “interactive-dialogue approach” and Van Dooren and Van de Walle’s [2008] “users”). Empirical evidence has offered some insights into the differences within stakeholder groups (e.g., Van Ryzin et al. 2004). However, stakeholders’ preferences for performance dimensions remain under-examined.

Performance data are drawn from stakeholders’ perceptions or archival sources (Selden and Sowa 2004; Walker, Boyne, and Brewer 2010). Perceptual measures are useful, as they cover a wide range of noneconomic dimensions of performance while being essential to core dimensions such as responsiveness. Perceptual measures are vulnerable to common methods bias problems, which is problematic when the measure of performance and its determinants are taken from a single dataset and analyzed (Jakobsen and Jensen 2015). Archival measures are considered more robust, as they are not subjective. However, they present validity concerns. They are socially constructed, as actors in the public realm make

decisions about what (and what not) to record as performance-related outputs or outcomes, and can be manipulated. Student test scores, which are frequently used in public management performance studies, usefully illustrate these tensions (Walker and Andrews 2015). The narrow focus on exam results does not account for the broad range of “value added” outputs and outcomes associated with effective schooling, such as developing character, integrity, and empathy for others: learning citizenship skills: and promoting social integration. Widespread cheating has also been found, which undermines the more objective status of archival measures (Andrews, Boyne, and Walker 2011).

Given these challenges, an ideal measure of performance that would comprehensively describe the achievements of an organization includes a range of performance dimensions, together with perceptual and archival data collected from stakeholders internal and external to the organization. This would ensure that appropriate PI measured in fitting ways is available to stakeholders. This is a somewhat unlikely expectation for research and practice. Rather, obtainable data must be tailored to the needs of stakeholders. Therefore, we focus on a smaller number of performance dimensions – efficiency, effectiveness and equity – and examine internal and external and archival and perceptual data types from the perspective of effectiveness.

Research Expectations in the Present Study

Service outcomes are important to internal and external stakeholders alike. Evidence from PIU studies has suggested that performance indicators and targets focusing on outcomes provide more valuable information than those focusing on outputs or activities (Boyne 2010). Building on these findings, we hone in on the longstanding social science debate about efficiency and equity, the origins of which lie in questions about the efficiency of capitalist economic institutions and the political principle of equality (Okun 1975). Arguments about efficiency are central to public choice criticisms of public services, which have in turn

informed the debate surrounding privatization and outsourcing (Boyne et al. 2003). Critiques of public choice have objected to these theorists' omission of the concept of equity, and have pointed to the role the state plays in addressing the inability of private markets to equitably allocate resources. We contrast these two dimensions of performance with goal achievement or specific organizational tasks or effectiveness. Effectiveness is the core objective of service delivery and ensures that organizational goals and tasks are met. This includes ensuring that children are educated and pass their examinations, or that solid waste is recycled. Research on managers has examined, through observational data, whether they prioritize particular performance indicators when delivering services: the evidence has suggested that they do not (Kelman and Friedman 2009; Mikkelsen 2017). For citizens, research findings have suggested that they do differentiate between performance indicators (Andrews and Van de Walle 2013; Grosso, Charbonneau and Van Ryzin 2017; Ho 2008).

We anticipate that priming people to consider different aspects of performance affects the way information about the effectiveness of a program is interpreted. In this way, performance is a more complex multidimensional concept than simply considering whether a program's goals are achieved and assessing how well an organization is doing, as it depends on how the information is contextualized. We develop a secondary anticipated relationship from this insight and hypothesize that those involved in the production of secondary education and recycling (internal stakeholders) prioritize the performance dimensions of effectiveness over efficiency and equity. Education is a high stakes service in which increased emphasis is placed on student grades – Hong Kong is not different in this regard. Environmental policy managers in the recycling sector also prioritize effectiveness. They have vested interests in raising recycling rates above the relatively low rates achieved in Hong Kong compared to other jurisdictions. We are unable to identify any systematic body of research that as specifically examined these issues among external stakeholders or citizens.

Andrews and Van de Walle (2013) showed that the level of disadvantage influenced citizens' assessments of performance while Ho's (2008) case study evaluation of citizen's preference for performance measures concludes that they select outcome measures. We, therefore, speculate that external stakeholders will prefer equity in education and in solid waste management.

Our second research expectation focuses on the type of complexity surrounding performance: the data sources and types. Academic debates have raged over the merits of perceptual and archival data. We hypothesize that internal and external stakeholders will hold preferences for performance data that are archival and external. The argument behind this expectation is one derived not from the nuances of statistical measurement issues, but accountability. Data on performance that are external and provided at some distance from an organization under scrutiny are likely to be perceived as offering a higher level of accountability, consistent with findings that external sources of information can boost its credibility for those being given the information (James and Van Ryzin 2017). External data can be perceptual or archival. Archival data has also been proposed as offering a higher level of accountability than perceptual data, given their distance from the organization being held to account, by being more objective and appearing harder to manipulate. However, we note the lack of a robust set of evidence on these questions from the perspective of PIU. We anticipate these expectations holding for internal and external stakeholders in the research settings of education and solid waste.

Research Design

Evidence gleaned from research carried out using experimental methods has proved to be fruitful in public administration research and practice, building upon the widespread use of experimental methods in associated disciplines (James, Gilke, and Van Ryzin 2017; John et al. 2011; Torgerson and Torgerson 2008). Experiments with randomization promise to isolate

cause-and-effect relationships by creating statistically equivalent groups that differ systematically only in whether they are subject to experimental treatment, eliminating the problems, such as omitted variable bias, that often plague multivariate regression and related analytical techniques (Cook, Campbell, and Day 1979). We implement an EVM which “consists of presenting participants with carefully constructed and realistic scenarios to assess dependent variables including intentions, attitudes, and behaviors ... and also allows researchers to manipulate and control independent variables, thereby simultaneously enhancing both internal and external validity” (Aguinis and Bradley 2014, 352). The careful construction of an EVM ensures the integrity of treatments, subjects, contexts and outcomes (Gerber and Green 2012; Harrison and List 2004).

Having said this, it is recognized that the EVM presents hypothetical scenarios to subjects that might not reflect the actual ways in which citizens use performance information to make judgements about public services. In this article, we focus on dimensions of performance and the source and type of data, yet emerging research in this field points towards other factors that influence evaluations of public service performance (Baekgaard 2015; Baekgaard and Serritzlew 2016; James and John 2007). To help address these limitations and to ensure a rigorous and transparent research design we follow best practice and reporting guidelines for social science experimental designs (Boutron et al. 2008; Geber et al. 2014).

Research setting

We apply EVM to two policy arenas in Hong Kong: secondary education and solid waste recycling. These two services were chosen for two reasons. First, secondary education is a person orientated service, and solid waste recycling is a technical service. Second, stakeholder groups have different roles and positions in relation to their production and consumption. Furthermore, each policy area is a high-profile topic in the Hong Kong. We

describe this below and elaborate further in our discussion of the vignettes below.

In secondary education, internal stakeholders include teachers, school managers and parents who have a direct role in and understanding of the delivery and production of education services. External stakeholders outside of the school system are also likely to have interest in secondary education as citizens and taxpayers may be concerned with the effectiveness of schooling. For example, youth English language skills are often challenged in Hong Kong's media. Secondary education is a critical deciding factor in students' career paths, and high-stakes examinations such as the Hong Kong Diploma of Secondary Education (HKDSE) are important. Expectations of students in secondary education is high, students face many pressures on their time inside and outside school. In school students experience long days with large amounts of homework and also engage in a wide range of extra-curricular activities including additional tutoring.

Solid waste is a pressing topic, as landfill sites in Hong Kong are rapidly filling and there is considerable disagreement over the technical feasibility and costs of alternative disposal options, such as recycling. Recycling rates are low in Hong Kong, compared to international benchmarks. While garbage bins are frequent in shopping malls and on the street, recycling facilities are few and far between. A shopping mall with many floors and shops may only have one or two recycling bins. In the recycling study, internal stakeholders are environmental managers and policy experts who are active in the solid waste and environmental arena. External stakeholders are citizens with lower levels of expertise than trained professionals. There are clear quantitative performance data available for both services provided by the Hong Kong Government and this data informed the design of the vignettes.¹

Participants and Procedures

Participants. 554 Hong Kong residents were recruited from My Citizens Panel as subjects.²

The Panel is hosted by the Laboratory for Public Management and Policy at City University of Hong Kong. The survey was open from September 2015 to May 2016. Under one third of the total respondents (n = 144, 32.5%) were experts, employed in either the education as secondary school teachers (n = 91, 16.43%) or worked in the environmental policy field (n = 89, 16.06%).³ Parents, some of the internal stakeholders used in the education experiments, formed just under 30% of the respondents (n = 157, 28.3%). The external stakeholders, for the solid waste experiments, included all of the respondents who were not environmental managers or policy experts (n = 453, 81.77%).⁴

The subjects' socio-economic characteristics varied from those of the Hong Kong population. The subjects were typically more educated than the overall population and were younger (38.63% of the respondents were in their 30s). The higher levels of education might have affected the subjects' interpretation of PI (Olsen 2016). The subjects were also e-literate thus might have been familiar with receiving large amounts of information.

Procedures. Pre-test questions were carefully arranged to prevent priming effects. The subjects were asked to answer questions about their general satisfaction with 10 diverse policy areas including secondary schools and household solid waste recycling. Personal demographic questions were also included. After completing the pre-test items, the subjects read vignettes.

A between-subjects design was adopted whereby each participant read one vignette from each experiment. Random allocation of the vignettes across the subjects was achieved using the functions in Qualtrics, the software used to deliver the study to the subjects. After reading the vignette the subjects were then asked to assess the performance of an organization and were asked: "Based on this story, how do you assess (organization name)'s (mathematics/solid waste recycling) performance?" The response scale for each question was a 7-point Likert scale, anchored with 1 = very negative and 7 = very positive. The survey

took 17 minutes to complete on average. Given its length, the subjects were compensated HK\$100 (US\$12.90) for their participation. Ethical approval was gained from XXXX (#H000854).

Experiment vignettes

We developed vignettes for two experiments in two important and topical policy areas in Hong Kong (see above). The vignettes and survey questions were piloted in and validated by focus groups of experts with detailed knowledge of secondary education and solid waste policy in Hong Kong.⁵ In order to ensure that performance assessments were not affected by the subjects' personal experiences, the vignettes used pseudonyms but drew on real performance data. Each vignette was positively framed, reporting high performance. For education, the vignettes focused on the mathematics performance of a secondary school in the 2014 HKDSE and drew performance data from the report. The environment vignette described recycling in a shopping mall, because all Hong Kong residents would have regular experience of going to a mall. The data in the vignettes were drawn from 2014 and taken from the Hong Kong Environmental Protection Department's website. (The full text of the vignettes is provided in the Appendix, Table A1 and Table A2.)

Experiment 1: Performance dimensions. Experiment 1 established contrasting vignettes to determine whether giving stakeholders primes about efficiency or equity affected how they interpreted the same information about program effectiveness. The measures of performance are outlined in table 1. Experiment 1 presented factual information about the effectiveness (examination score or ranking of the school/recycling rates) of the program as a baseline, such that it was possible to measure differences from the achievement of formal goals. This was delivered in treatment C.

[Table 1 here]

Treatment group A subjects received additional information about efficiency. The

school vignette focused on the timely and thus efficient delivery of the curriculum given the education context in Hong Kong. The vignette states “Lee Ka Ming Memorial School does this by reducing the curriculum time while keeping the high score-graduation rate”. In Hong Kong, the education system is highly selective, with selection based on the results of public examinations within the context of a uniform curriculum for all (Pong and Chow 2002). Being able to complete the uniform curriculum efficiently is important because students are under pressure to perform well, have many extra-curricular activities, and tutorials and are expected to achieve high scores. Time and its efficient use is thus critical in Hong Kong and students, and teachers are very time sensitive. The solid waste vignette presented subjects with the following statement: “The shopping center managers have tried to cut recycling costs in multiple ways while maintaining a good recycling rate.” This taps into the notion of efficiency as cost per unit of output. The statement reflects the efficient way that the malls are managed in general, and in particular, the cost pressures managers face on a daily basis including management practices such as recycling.

Treatment group B focused on the access criteria of the performance dimension equity. Subjects were provided with the following statement on equity: “Lee Ka Ming Memorial School takes an equity focus to boost school’s mathematics performance. Lee Ka Ming Memorial School does this by offering extra support for low scored-students to make sure that no student drops out of mathematics”. As noted above, Hong Kong education system focuses heavily on a uniform system that may not suit all students (Pong and Chow 2002), and yet being able to meet the public examination requirements is the only way to seek further education for most students. A school that seeks to maintain the majority of students in these high stakes tests, including the provision of additional and different resources (extra tutoring in the above example) to ensure all students receive what is needed to meet performance targets is in line with the definition of educational equity, and the notion

of fair access.

For the solid waste recycling vignette subjects were primed with the following statement: “Lingfan Garden Plaza shopping center focuses on achieving an equitable access recycling program. The shopping center managers provide a similar number of recycling bins on every floor.” Hong Kong is a dense high rise city. For example, fifty percent of people live in public housing estates that have more than 10 high-rise buildings of 35-40 stories with more than 10 family units on each floor: each unit is around 40m². Each estate has its own shopping mall that residents use for all their meals and daily shopping, and are thus very familiar with. In these very crowded places there are many garbage cans, as often as every 30 steps. However, there are few recycling bins, maybe at the concierge desk. In these dense environments, someone might not see any recycling bins in their walking routes. The equitable distributions of recycling bins, for example, on each floor of a shopping mall, makes them much more accessible to Hong Kong residents.

Experiment 2: Performance data sources and types. The second experiment examined data types and data sources. The experiment was a two-by-two factorial design.⁶ Each vignette built upon the context in Hong Kong for each topic and again primed subjects with information on effectiveness, highlighting the attainment of organizational goals. Figure 2 provides information on the source of the PI for the four treatments groups.

[Figure 2 here]

The external data source were external reviewers in the external-perceptual and the external-archival vignettes. For example, in the education experiment subjects received the following stimuli “An Education Expert Team, composed of professors from local teaching universities, was asked to develop a Score Card ...”. The internal data source vignettes stated the internal source as “The School Management Board of Victoria College ...”. For the perceptual data vignettes described the school performance as “... an excellent example to

other schools in Hong Kong”. The archival vignettes described the school performance “... among the top 5% of best performing schools for the percent of students attaining the best grade at mathematic compulsory in 2014 in the HKDSE”.

The text for the solid waste recycling vignettes have same structure. The internal data source is the shopping center managers’ performance report and the external data source is a nongovernmental organization named “Environmental Green”. The perceptual data vignettes presented subjects with the following statement: “[managers/Environmental Green] ... have reported to other shopping malls their experiences and recommended their own approach as an excellent example to other shopping malls”. The archival data reported to subject that “... Ocean Plaza achieved a 46% recycling rate”.

Analysis

The data analysis was based on difference-in-means tests and regressions. The results tables provide the mean score for each vignette together with standard deviations. The ANOVA results are presented and statistical differences between response categories are noted. Ordinal logit regressions were also conducted and analysis undertaken in STATA. Dummy variables were used to record the stakeholders – internal stakeholders with external stakeholders as the referent case. Alongside these stakeholder groups, pre-test measures were included for demographics: gender, length of residence in Hong Kong (more than 10 years) and age. One further item was included from the pre-test to control for priming effects: a question about the assessment of performance of secondary schools and solid waste recycling services in Hong Kong. The regression results confirmed the findings from the ANOVA analysis and are and be found in the appendix, tables A1-A2. In a number of cases, the controls were statistically significant, these are reported.

Results

Experiment 1

Experiment 1 examined the performance dimensions of efficiency, equity and effectiveness. Table 2 presents descriptive data by internal and external stakeholders for education and recycling and ANOVA results by experiment treatments (efficiency, equity, and effectiveness) and stakeholder groups.

[Table 2 here]

For the education vignettes, both the internal and external stakeholders gave different scores for the organization's performance when given information about effectiveness in the context of discussing equity, compared with when effectiveness was given alone and in the context of efficiency (left-upper hand panel of table 2). The ANOVA showed a main vignette effect: a statistically significant difference in performance ratings for treatments (bottom-left hand panel of table 2). The post-hoc analysis showed that the participants who read the equity and effectiveness (by itself) vignettes demonstrated higher satisfaction with the organization than those who read the efficiency vignette. Notably, internal stakeholders rated effectiveness (by itself) above both efficiency and equity, whereas the external stakeholders offered the greatest support for equity. However, the interaction between stakeholders and treatments was not statistically significant. The logit regression analysis (table A3) indicated that one control was statistically significant and of interest. The subjects who demonstrated higher levels of satisfaction with secondary schooling in the pre-test rated performance higher when scoring the school math vignettes. This suggests that the subjects' inherent predispositions toward secondary school education are reflected in the results.

For the solid waste recycling vignettes, the group that received additional information about equity also gave high rankings, but differences across the treatments were not statistically significant (upper-right panel of table 2). The external stakeholders offered more positive assessments of all of the performance vignettes, such that the differences between internal and external stakeholders were statistically significant (bottom-left hand panel of

table 2), suggesting that internal stakeholders are perhaps more skeptical about recycling performance. As with the education results, the interaction between stakeholders and treatments was not statistically significant.

Experiment 2

Experiment 2 tested the effects of data sources (internal and external source) and data type (perceptual and archival data) as shown in table 3. From the left side, external-perceptual data, internal-perceptual data, external archival data, and internal-archival data are displayed showing variations in assessment of the data sources assessment.

[Table 3 here]

The descriptive data and ANOVA results of the secondary education experiment are presented in the left-hand panel of table 3, showing a number of statistically significant differences. The results indicate that external-archival data are preferred over the internal ones in the mathematics secondary education vignettes (left panel of table 3). The post-hoc analysis indicated that all of these differences are statistically significant. Similar to the previous experiments, the regression results showed that positive previous ratings of secondary education influenced the results (table A4). The differences between the stakeholders and the interaction between stakeholders and treatments were not statistically significant.

The results of the solid waste recycling experiment (right panel of table 5) also showed statistically significant differences, but the level of significance was less than in the secondary education experiment. The external-archival data were rated the highest. In the post-hoc analysis, all of the statistically significant differences were with this data type, including internal-perceptual data, and internal-archival data. The differences between the stakeholders and for the interaction between stakeholders and treatments were not statistically significant. The logit regression results (table A4) indicated that previous attitudes towards

recycling are likely to have a positive association with the subjects' assessment of performance in the vignette, as is subject age.

Discussion

We sought to unpack PIU and the complex concept of performance. We did this by assessing different dimensions of performance and forms of PI across internal and external stakeholders and services, including secondary education and solid waste recycling. We examined subjects' attitudes towards dimensions of performance (efficiency, effectiveness and equity), and data sources (internal, external, perceptual and archival) in three experiments, respectively.

The findings point toward important subtleties in the attitudes of stakeholders in different services. We found that stakeholder assessments of performance dimensions vary by service area. Effectiveness information alone and the provision of further information about equity led to higher performance ratings than when further information about efficiency was provided in secondary education. There was no statistically significant difference between the provision of these different forms of information in the case of solid waste recycling. However, there were differences between the stakeholders' assessments in both experiments. Secondary education internal stakeholders prioritized effectiveness, whereas external stakeholders ranked equity higher. In the recycling experiment, the external stakeholders consistently offered higher performance assessments, with the internal stakeholders offering the lowest assessments of effectiveness. The secondary education experiment offered clear support to the hypothesis that internal stakeholders favor effectiveness and external stakeholders do equity. However, the results were not as robust for recycling. Findings that the provision of PI affects the attitudes of stakeholders' attitudes towards and assessments of services is widely noted in the literature, but is typically drawn from a single stakeholder group (James and Moseley 2014; Van Ryzin et al. 2004). Baekgaard's (2015) PIU study of

users and citizens, set in secondary education, showed that PI (effectiveness data) can influence user's attitudes towards their services, and even more so when supplemented with cost data. Stakeholders' preferences for different performance dimensions may have important influences on how assessments of service performance are made, such that the most pertinent dimension of performance must target relevant groups. This contention must be examined and replicated in different contexts.

All of the vignettes in Experiment 2, excluding those with internal-perceptual data had positive effects on performance. This suggests that the provision of internal-perceptual data is not well received by the stakeholder groups in this article and that the data type with the strongest effect on performance assessments is external-archival data. These findings were consistent across the stakeholder groups and reflect the hypothesized expectation of our discussion on the sources of performance data. These findings also connect to ongoing debates in public management research about the use of archival and perceptual data (Jakobsen and Jensen 2015). Arguments have been made that perceptual and self-reported data suffer from common source bias. Academics have argued that these data should not be used as dependent variables in studies of performance. Our evidence supports this contention, as experts and citizens quizzed in this experiment rated internally generated perceptual data on par with a vignette that contained no PI, and rated external-archival data most favorably.

Our findings reinforce our argument that performance is a complex construct. The results point toward important nuances between services and stakeholders that have conceptual and practical implications for the study of performance and its use. Our examination of attitudes and preferences of data sources and types also contributes to ongoing debates, more recently seen in "behavioral public administration" which draws upon psychology. Baekgaard and Serritzlew (2015) and James and Van Ryzin (2016) noted how motivated reasoning can influence stakeholder's assessment of performance and Olsen

(2016) discussed the important role episodic information played in performance assessments. These influences were found in our secondary education experiments. The regression models identified that subjects with previously favorable dispositions towards secondary education also offered higher assessment of performance in the vignettes, whereas this was noted in only one instance in the recycling study.

These findings also have implications for government and PIU practices. Public service organizations should consider providing more targeted and appropriate PI to stakeholders: appropriate dimensions of performance must be identified, and emphasis should be placed on the provision of relative PI and archival and externally generated data. Governments must also be aware that stakeholders' predispositions play an important role in their performance assessments (Baekgaard and Serritzlew 2015; James and Van Ryzin 2016; Olsen 2016).

A number of limitations that weakened the generalizability of its results, thereby offering opportunities for further research are noted. First, the study was conducted in Hong Kong. Sampling from this population is beneficial in providing findings beyond the U.S. or European contexts within which most studies have been conducted. However, the results may be applicable only to this jurisdiction, with its particular characteristics that were outlined when discussing research design and the vignettes. Second, the results may differ for other stakeholder groups and alternative policy arenas. Efficiency, effectiveness and equity were examined in Experiment 1. Subjects were faced with hypothetical vignettes that manipulated effectiveness and efficiency and effectiveness and equity yet when internal or external stakeholders face real world performance they will contain many alternative performance dimensions which may reveal different priorities. The performance constructs could also be operationalized in alternative ways. For example, a measure of efficiency that examined the relationship between inputs and outputs would tap into a more traditional operationalization

of the construct. In the same vein, Experiment 2 presented four combinations of internal/external and perceptual/archival performance data. In practice, multiple data types and sources may be available providing internal and external stakeholders with many reference points to make a more holistic assessment of organizational achievements. Nonetheless, in building on the growing number of studies on PIU, by including two stakeholder groups and two policy areas we have made a small advancement in knowledge. Third, replication of this study, together with extensions in different populations using alternative measures and analyses may address some of these limitations. Finally, the experimental design offered strong insights to causality. Although we note effects at the time of the study, whether these effects are enduring is uncertain. Longitudinal studies would tease out these effects over time.

Conclusion

PIU has been one of the more persistent reforms to have emerged from the New Public Management some four decades ago. Despite its long history, new attitudinal studies, such as this one, and the more recent behavioral approach from the psychological perspective (Grimmelikhuisen et al 2016) indicate that much still needs to be uncovered on this topic. This research is important as it informs the PI communicated to key stakeholders, such as managers, experts, users and citizens, in addition to public management practices. Inappropriate PI is given to stakeholder groups has significant consequences for the provision of public services. We encourage others to continue studying PIU to develop knowledge on this key public management issue.

Notes

1. Performance data were taken from:

https://www.wastereduction.gov.hk/en/assistancewizard/waste_red_sat.htm;

<http://edb.gov.hk/en/sch-admin/sch-quality-assurance/performance-indicators/index.html>

2. We follow Mutz and Pemantle (2015) advice because we are not making claims of full representativeness of the sample to the Hong Kong population.
3. The designation of “expert” was given by the research team based upon questions in the pre-test survey (not reported here).
4. Sample sizes of the secondary education study were 541 for experiment 1, and 545 for experiment 3. Sample sizes across the three experimental studies in solid waste recycling are 532 and 531.
5. Prior to conducting the experiments, focus group interviews were held to discuss the resources to be used in the studies. The meetings were held between March 2015 and August 2015. The focus group members were experts in experimental research designs, education, and environmental policy and were drawn from XXXX. The focus groups discussion examined vignette content, sample characteristics, and the survey delivery strategy.
6. The original design of experiment #2 included a control. However, following a line of enquiry from one reviewer the control has been dropped. This has the advantage that the research design of the two studies is similar, but clearly brings a weakness in that there is no counterfactual. This change did not affect the general balance of the results (full results available on request from the authors).

References

Aguinis, Herman, and Kyle J. Bradley. 2014. Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational*

Research Methods, 17(4): 351-371.

Andersen, Lotte Bøgh, Andreas Boesen, and Lene Holm Pedersen. 2016. Performance in public organizations: Clarifying the conceptual space. *Public Administration Review*, 76(6): 852-862.

Andrews, Rhys, George Boyne, and Richard M. Walker. 2011. The impact of management on administrative and survey measures of organizational performance. *Public Management Review*, 13(2): 227-255.

Andrews, Rhys, and Steven Van de Walle. 2013. New public management and citizens' perceptions of local service efficiency, responsiveness, equity and effectiveness. *Public Management Review*, 15(5): 762-783.

Arnett, Jeffrey. 2008 The neglected 95%: Why American psychology needs to become less American. *American Psychologist* 63(7):602-14.

Baekgaard, Martin. 2015. Performance information and citizen service attitudes: Do cost information and service use affect the relationship? *International Public Management Journal*, 18(2), 228-245.

Baekgaard, Martin, and Søren Serritzlew. 2016. Interpreting performance information: Motivated reasoning or unbiased comprehension. *Public Administration Review*, 76(1): 73-82.

Behn, Robert D. 2003. Why measure performance? Different purposes require different measures. *Public Administration Review* 63(3): 586-606.

Boutron, Isabelle, David Moher, Douglas G. Altman, Kenneth F. Schulz, and Philippe Ravaud. 2008. Extending the CONSORT statement to randomized trials of

- nonpharmacologic treatment: explanation and elaboration. *Annals of Internal Medicine*, 148(4): 295-309.
- Boyne, George A. 2002. Concepts and indicators of local authority performance: an evaluation of the statutory frameworks in England and Wales. *Public Money and Management*, 22(2): 17-24.
- Boyne, George A. 2010. Performance management: Does it work? In *Public management and performance: Research directions*, Richard M. Walker, George .A. Boyne, and Gene A. Brewer, eds. Cambridge: Cambridge University Press.
- Boyne, George A., Catherine Farrell, Jennifer Law, Martin Powell and Richard M. Walker. 2003. *Evaluating public management reforms: Principles and practice*. Buckingham: The Open University Press.
- Cook, Thomas D., Donald T. Campbell, and Arles Day. 1979. *Quasi-experimentation: Design and analysis issues for field settings* (Vol. 351). Boston: Houghton Mifflin.
- George, Bert, Sebastian Desmidt, Poul A. Nielsen, and Martin Baekgaard. 2016. Rational planning and politicians' preferences for spending and reform: replication and extension of a survey experiment. *Public Management Review*
<http://dx.doi.org/10.1080/14719037.2016.1210905>
- Gerber, Alan S., Kevin Arceneaux, Cheryl Boudreau, Conor Dowling, Sunshine Hillygus, Thomas Palfrey, Daniel R. Biggers, and David J. Hendry. 2014. Reporting guidelines for experimental research: A report from the experimental research section standards committee. *Journal of Experimental Political Science* 1(1): 81-98.
- Gerber, Alan S., and Donald P. Green. 2012. *Field experiments: Design, analysis, and interpretation*. New York: WW Norton & Company.

- Grosso, Ashley, Étienne Charbonneau, and Greg G. Van Ryzin. 2017. How Citizens Respond to Outputs, Outcomes, and Costs: A Survey Experiment About an HIV/AIDS Program. *International Public Management Journal*, 20(1), 160-181.
- Grimmelikhuijsen, Stephan. G., Sebastian Gilke, Asmus L. Olsen, and Lars Tummers. 2016. Behavioral Public Administration: combining Insights from Public Administration and Psychology. *Public Administration Review*, 77(1): 45-56.
- Harrison Glenn W., and John A. List. 2004. Field experiments. *Journal of Economic Literature*, 42(4): 1009-55.
- Heinrich, Carolyn J., and Laurence E. Lynn. Jr. 2000. *Governance and performance: New perspectives*. Washington, DC.: Georgetown University Press.
- Holzer, Marc, and Kathryn Kloby. 2005. Public performance measurement: An assessment of the state-of-the-art and models for citizen participation. *International Journal of Productivity and Performance Management* 54(7): 517-532.
- Ho, Alfred Tat-Kei. 2008. Reporting public performance information: The promise and challenges of citizen involvement. In Van Dooren, Wouter and Van de Walle, Steven. (Eds.) *Performance information in the public sector: how it is used*. Basingstoke: Palgrave McMillan
- Jakobsen, Morten, and Rasmus Jensen. 2015. Common method bias in public management studies. *International Public Management Journal* 18(1): 3-30.
- James, Oliver. 2011. Performance measures and democracy: Information effects on citizens in field and laboratory experiments. *Journal of Public Administration Research and Theory* 21(2): 399-418.

- James, Oliver, Sebastian Jilke, and Gregg G. Van Ryzin. 2017. Causal Inference and the Design and Analysis of Experiments in James, O., Jilke, S. and Van Ryzin, G.G. (eds) *Experiments in Public Management Research: Challenges and Contributions*. Cambridge: Cambridge University Press.
- James, Oliver, and Peter John. 2007. Public management at the ballot box: Performance information and electoral support for incumbent English local governments. *Journal of Public Administration Research and Theory*, 17(4): 567-580.
- James, Oliver, and Alice Moseley. 2014. Does performance information about public services affect citizens' perceptions, satisfaction, and voice behaviour? Field experiments with absolute and relative performance information. *Public Administration*, 92(2): 493-511.
- James, O. and Van Ryzin, G.G., 2016. Motivated reasoning about public performance: An experimental study of how citizens judge the affordable care act. *Journal of Public Administration Research and Theory*, 27(1): 197-209.
- James, Oliver, and Gregg G. Van Ryzin. 2017. Incredibly good performance: An experimental study of source and level effects on the credibility of government. *The American Review of Public Administration* 47(1): .23-35.
- John Peter, Cotterill, Sarah, Liz Richardson Alice Moseley, Graham Smith, Gerry Stoker, and Corinne Wales. 2011. *Nudge, Nudge, Think, Think: Using Experiments to Change Civic Behaviour*. London: Bloomsbury.
- Kelman, Steven. and John N. Friedman. 2009. Performance improvement and performance dysfunction: An empirical examination of distortionary impacts of the emergency

- room wait-time target in the English National Health Service. *Journal of Public Administration Research and Theory* 19(4): 917-46.
- Machery, Edouard. 2010. Explaining why experimental behavior varies across cultures: a missing step in “the weirdest people in the world?” *Behavioral and Brain Sciences* 33(2-3): 101-102.
- Mikkelsen, Maria F. 2017. Do managers face a performance trade-off? Correlations between product and process performance. *International Public Management Journal* doi.org/10.1080/10967494.2016.1276495.
- Mitchell, Ronald K., Bradley R. Agle, and Donna J. Wood. 1997. Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of Management Review* 22(4): 853-886.
- Moynihan, Donald P. 2008. *The dynamics of performance management: Constructing information and reform*. Washington, DC: Georgetown University Press.
- Mutz, Diana C. and Robin Pemantle. 2015. Standards for Experimental Research: Encouraging a Better Understanding of Experimental Methods. *Journal of Experimental Political Science*, 2(2): 192-215.
- OCED. 2005. *Moderizing government: The way forward*. Paris: OECD.
- Okun, Arthur. 1975. *Efficiency and equity: The big tradeoff*. Washington, DC: The Brookings Institute.
- Olsen, Asmus L. 2016. Human Interest or Hard Numbers? Experiments on Citizens’ Selection, Exposure, and Recall of Performance Information. *Public Administration Review* 10.1111/puar.12638

- Pettigrew, Andrew, Ewan Ferlie, and Lorna McKee. 1992. Shaping strategic change-The case of the NHS in the 1980s. *Public Money & Management*, 12(3): 27-31.
- Pollitt, Christopher. 2006. Performance information for democracy: The missing link? *Evaluation* 12(1): 38-55.
- Selden, Sally C., and Jessica E. Sowa. 2004. Testing a multi-dimensional model of organizational performance: Prospects and problems. *Journal of Public Administration Research and Theory* 14(3): 395-416.
- Sharpe, Lawrence J. 1970. Theories and values of local government. *Political Studies* 18(2):153–174.
- Thompson, James D. 1967. *Organizations in action: Social science bases of administrative theory*. New York: McGraw-Hill.
- Torgerson David, and Carole Torgerson. 2008. *Designing randomised trials in health, education and the social sciences*. Basingstoke: Palgrave Macmillan.
- Van Dooren, Wouter and Steven Van de Walle. (Eds.) 2008. *Performance information in the public sector: How it is used*. Basingstoke: Palgrave McMillan.
- Van Ryzin, Gregg G., Douglas Muzzio, Stephen Immarwahr, Lisa Gulick, and Eve Martinez. 2004. Drivers and consequences of citizen satisfaction: An application of the American Customer Satisfaction Index model to New York City. *Public Administration Review*, 64(3): 331-341.
- Walker, Richard M., and Rhys Andrews. 2015. Local government management and performance: a review of evidence. *Journal of Public Administration Research and Theory* 25(1): 101-133.

Walker, Richard M., George A. Boyne, and Gene A. Brewer. 2010. *Public management and performance: Research directions*. Cambridge: Cambridge University Press.

Figure 1 Stakeholders and data sources in public service performance: Examples of effectiveness in secondary school education

		Stakeholders	
		Internal	External
Types of data	Perceptual	<p>Staff survey e.g., staff perceptions of student attainment</p>	<p>Citizen survey e.g., citizen perceptions of student attainment</p>
	Archival	<p>Internal data e.g., class test results held by teacher</p>	<p>Audited performance indicators e.g., formal school test results</p>

Figure 2 Performance data type in vignettes

		Data source	
		Internal	External
Data	Perceptual	Education	Treatment A Secondary school management board's performance review
		Environment	Treatment B Shopping center managers' performance review
	Archival	Education	Treatment C Secondary school management board's performance score card
		Environment	Treatment D Shopping center managers' performance report

Table 1 Illustration of types of performance information

	Treatment A: Efficiency	Treatment B: Equity
Education	Efficient delivery of the student curriculum	Extra time supporting on low scored-students in the curriculum
Waste	Cost of recycling per tonne	Access to recycling resources across the community

Table 2 Descriptive data and ANOVA results: experiment 1

<i>Descriptives</i>		Education			Solid Waste		
		i Efficiency	ii Equity	iii Effectiveness	i Efficiency	ii Equity	iii Effective ness
Internal Stakeholders	Mean	5.03	5.19	5.29	4.62	4.8	4.4
External Stakeholders	S.D.	1.28	1.05	1.09	1.18	1.47	1.4
Internal Stakeholders	Mean	4.75	5.25	5.01	4.87	4.97	4.96
External Stakeholders	S.D.	1.1	1.17	1.2	1.1	1.01	1
<i>ANOVA</i>		df	F	Partial Eta Squared	df	F	Partial Eta Squared
Treatments (T)		2	3.953**	0.015	2	0.928	0.004
Stakeholders (S)		1	2.659*	0.005	1	6.631***	0.012
T*S		2	1.139	0.004	2	0.885	0.003
<i>Post hoc tests</i>			i-ii*** i-iii* Internal-external*		Internal-external***		

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 3 Descriptive data and ANOVA results: experiment 2

<i>Descriptives</i>		Education				Solid Waste			
		i	ii	iii	iv	i	ii	iii	iv
Internal Stakeholders	Mean	4.8	4.59	5.27	4.9	4.67	4.62	5.13	4.5
External Stakeholders	S.D.	0.92	1.06	0.79	0.94	1.33	0.96	1.41	1.29
Internal Stakeholders	Mean	4.78	4.38	5.22	4.9	5.09	4.54	5.11	4.98
External Stakeholders	S.D.	1.17	1.2	0.98	1.05	1.01	1.13	0.96	1.03
<i>ANOVA</i>		df	F	Partial Eta Squared		df	F	Partial Eta Squared	
Treatments (T)		3	9.538***	0.060		3	2.427*	0.018	
Stakeholders (S)		1	0.510	0.001		1	2.034	0.005	
T*S		3	0.226	0.002		3	1.099	0.008	
<i>Post hoc tests</i>			i-ii**					ii-iii***	
			i-iii***					iii-iv**	
			ii-iii***						
			ii-iv***						
			iii-iv**						

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Key

i: External-Perceptual, ii: Internal-Perceptual, iii: External-Archival, iv: Internal-Archival

Appendix:

Table A1 Experiment 1 vignettes for performance dimension

Education

Treatment A: Efficiency

Education is a hot issue in Hong Kong. The Education Bureau plans to evaluate school performance to strategically develop a long-term education plan. Parents are very interested in school performance. Lee Ka Ming Memorial School takes an efficiency focus to boost school's mathematics performance. Lee Ka Ming Memorial School does this by reducing the curriculum time while keeping the high score-graduation rate. It seems that Lee Ka Ming Memorial School's efficiency focused policy worked well: Lee Ka Ming Memorial School achieved top 5% of best performing schools for the percent of students attaining the best grade at mathematics in 2014 HKDSE.

Treatment B: Equity

Education is a hot issue in Hong Kong. The Education Bureau plans to evaluate school performance to strategically develop a long term education plan. Parents are very interested in school performance. Lee Ka Ming Memorial School takes an equity focus to boost school's mathematics performance. Lee Ka Ming Memorial School does this by offering extra supporting for low scored-students to make sure that no student drops out of mathematics. It seems that Lee Ka Ming Memorial School's equity focused policy worked well: Lee Ka Ming Memorial School achieved top 5% of best performing schools for the percent of students attaining the best grade at mathematics in 2014 HKDSE.

Treatment C: Effectiveness

Education is a hot issue in Hong Kong. The Education Bureau plans to evaluate school performance to strategically develop a long term education plan. Parents are very interested in school performance. Lee Ka Ming Memorial School takes an education performance focus to boost school's mathematics performance. Lee Ka Ming Memorial School does this by managing time and students in their curriculum. It seems that Lee Ka Ming Memorial School's focus on education performance policy worked well: Lee Ka Ming Memorial School achieved top 5% of best performing schools for the percent of students attaining the best grade at mathematics in 2014 HKDSE.

Recycling

Treatment A: Efficiency

In densely populated Hong Kong solid municipal waste is a problematic issue as the city's landfills become full. The Environmental Protection Department plans to prevent and reduce all types of waste at source, and to increase reuse and recycling, as laid out in the Action Blueprint. According to recent Waste Statistics (2013) municipal solid waste generation in Hong Kong is around 9,457 tonnes per day. Lingfan Garden Plaza shopping center focuses on developing an efficient recycling program. The shopping center managers have tried to cut recycling costs in multiple ways while maintaining a good recycling rate. It seems that the efficiency policy has worked well; Lingfan Garden Plaza achieved a 46% recycling rate in 2014.

Treatment B: Equity

In densely populated Hong Kong solid municipal waste is a problematic issue as the city's landfills become full. The Environmental Protection Department plans to prevent and reduce all types of waste at source, and to increase reuse and recycling, as laid out in the Action Blueprint. According to the recent Waste Statistics (2013) municipal solid waste generation in Hong Kong is around 9,457 tonnes per day. Lingfan Garden Plaza shopping center focuses on achieving an equitable access recycling program. The shopping center managers provide a similar number of recycling bins on every floor. It seems that this equity policy has worked well; Lingfan Garden Plaza achieved a 46% recycling rate in 2014.

Treatment C: Effectiveness

In densely populated Hong Kong solid municipal waste is a problematic issue as the city's landfills become full. The Environmental Protection Department plans to prevent and reduce all types of waste at source, and to increase reuse and recycling, as laid out in the Action Blueprint. According to the recent Waste Statistics (2013) municipal solid waste generation in Hong Kong is around 9,457 tonnes per day. Lingfan Garden Plaza shopping center focuses on developing a recycling program. The shopping center managers try to achieve a good recycling rate. It seems that their recycling program has worked well; Lingfan Garden Plaza achieved a 46% recycling rate in 2014.

Table A2 Experiment 2 vignettes for data sources and types

Education

Treatment A: External-Perceptual data

An Education Expert Team, composed of professors from local teaching universities, was asked to develop a Score Card for each secondary school in Hong Kong. The Score Card for Victoria College noted the way that the Principal and teachers emphasizes the school's mathematics curriculum and the way they focus their efforts on managing time and students. The Education Expert Team reviewed Victoria College's mathematics curriculum as an excellent example to other schools in Hong Kong.

Treatment B: Internal-Perceptual data

The School Management Board of Victoria College reviewed their secondary school performance, and produced a Score Card. The Score Card for Victoria College noted the way that the Principal and teachers emphasizes the school's mathematics curriculum and the way they focus their efforts on managing time and students. The School Management Board reviewed Victoria College's mathematics curriculum as an excellent example to other schools in Hong Kong.

Treatment C: External-Archival data

An Education Expert Team, composed of professors from local teaching universities, was asked to develop a Score Card for secondary schools in Hong Kong. The Score Card for Victoria College emphasizes the school's mathematics curriculum and the way they focus their efforts on managing time and students. The Education Expert Team reported Victoria College's performance, noting that its achievements were among the top 5% of best performing schools for the percent of students attaining the best grade at mathematic compulsory in 2014 in the HKDSE.

Treatment D: Internal-Archival data

The School Management Board of Victoria College reviewed their secondary school performance, and produced a Score Card. The Score Card for Victoria College emphasizes the school's mathematics curriculum and the way they focus their efforts on managing time and students. The School Management Board Score Card reported Victoria College's performance for internal school tests. The Score Card noted that students typically achieved amongst the top 5% mathematics performance of schools in Hong Kong for internal school tests.

Recycling

Treatment A: External-Perception data

Because of the continued growth in municipal solid waste Hong Kong faces running out of landfill space in the next three years. The need to control municipal solid waste is increasingly important to fulfill environment and sustainable development targets. The Environmental Protection Department plans to prevent and reduce all types of waste at source, and increase reuse and recycling, as laid out in the Action Blueprint. Ocean Plaza Shopping Centre focuses on their recycling program and the shopping centre managers try to achieve a good recycling rate. It seems that their recycling program has worked well: the nongovernmental organization Environmental Green reviewed Ocean Plaza's recycling program and recommend its practices as an excellent example to other

shopping malls.

Treatment B: Internal-Perception data

Because of the continued growth in municipal solid waste Hong Kong faces running out of landfill space in the next three years. The need to control municipal solid waste is increasingly important to fulfill environment and sustainable development targets. The Environmental Protection Department plans to prevent and reduce all types of waste at source, and increase reuse and recycling, as laid out in the Action Blueprint. Ocean Plaza Shopping Centre focuses on their recycling program and the shopping centre managers try to achieve a good recycling rate. It seems that their recycling program worked well: Ocean Plaza managers have reported to other shopping malls their experiences and recommended their own approach as an excellent example to other shopping malls.

Treatment C: External-Archival data

Because of the continued growth in municipal solid waste Hong Kong faces running out of landfill space in the next three years. The need to control municipal solid waste is increasingly important to fulfill environment and sustainable development targets. The Environmental Protection Department plans to prevent and reduce all types of waste at source, and increase reuse and recycling, as laid out in the Action Blueprint. Ocean Plaza Shopping Centre focuses on their recycling program and the shopping centre managers try to achieve a good recycling rate. It seems that their recycling program worked as well: the nongovernmental organization Environmental Green reviewed their recycling performance in 2014 and noted that Ocean Plaza achieved a 46% recycling rate.

Treatment D: Internal-Archival data

Because of the continued growth in municipal solid waste Hong Kong faces running out of landfill space in the next three years. The need to control municipal solid waste is increasingly important to fulfill environment and sustainable development targets. The Environmental Protection Department plans to prevent and reduce all types of waste at source, and increase reuse and recycling, as laid out in the Action Blueprint. Ocean Plaza Shopping Centre focuses on their recycling program and the shopping centre managers try to achieve a good recycling rate. It seems that their recycling program worked as well: based on the data collected by managers over 2014 Ocean Plaza reported that they achieved a 46% recycling rate.

Table A3 Regression results: experiment 1

<i>Regression</i>	Education		Solid Waste	
	Coef.	z	Coef.	z
Efficiency ^a	-0.41	-2.14**	-0.1	-0.53
Equity	0.20	1.01	0.01	0.06
internal stakeholders	0.17	0.96	-0.46	-2.07**
male	0.14	0.86	0.07	0.43
HK	-0.02	-0.09	-0.33	-1.54
age	0.06	0.67	0.04	0.5
Satisfaction in HK education/SW policy	0.01	2.23***	0.00	0.18
N	529		521	
Prob>chi2	0.0091		0.3143	
Pseudo R2	0.0115		0.0054	

** $p < 0.05$; *** $p < 0.01$

^a Effectiveness is excluded

Table A4 Regression results: experiment 2

	Education		Solid Waste	
	Coef.	z	Coef.	z
External-perceptual ^a	0.61	2.47**	1.04	3.87***
External-archival	1.44	5.67***	1.19	4.35***
Internal-archival	0.76	3.04***	0.77	2.89***
Internal stakeholders	-0.03	-0.14	-0.32	-1.28
Male	0.15	0.86	0.33	1.77
HK	0.18	0.77	-0.12	-0.45
Age	0.04	0.45	0.17	1.89*
Satisfaction in HK education/SW policy	0.01	3.32***	0.00	-0.19
N	447		402	
Prob>chi2	0.000		0.000	
Pseudo R2	0.0353		0.0286	

* $p < 0.1$; ** $p < 0.05$;
*** $p < 0.01$

^a Internal-perceptual is excluded