Measurement and benchmarking foundations: Providing support to organizations in their development and growth using dashboards

Devender Maheshwari *, Marijn Janssen

Delft University of Technology, Jaffalaan 5, 2628 BX Delft, The Netherlands

ARTICLE INFO

ABSTRACT

Growth and stage models often lack a sound empirical and theoretical base and do not provide any help for organizations to improve. Measuring and benchmarking (M&B) is necessary for understanding an organization's position and identifying growth opportunities. Yet M&B methods are often not based on generalizations of practice and measure only what is directly visible. They are missing relevant elements that can help further development. In this paper, we propose a multi-level measurement framework utilizing a mix of measurement methods to look deep inside organizations. Whereas benchmarking is often based on a single number, deep insight is given by showing the performance in a broad range of areas and views using a dashboard. Guidance for improvement is created by identifying those elements that need improvements. The illustration of the framework in a case study shows that the process of measuring deep inside organizations might be more important than the actual outcomes and that per area different maturity levels might be possible. We provide seven principles that can serve as a foundation for developing M&B and stage models.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Measurement and benchmarking (M&B) of public organizations is becoming increasingly important for governments and organizations to help them in their development and to take advantage of the newest developments. Benchmarking is the measurement of some elements and the comparison of the outcomes to a certain norm, the benchmark. Whereas the focus of benchmarking is on mutual learning, practice often shows a narrow focus on numbers (Bannister, 2007), most models are developed for certain situations (Coursey & Norris, 2008), and success has been limited so far (Ojo, Janowski, & Estevez, 2011). Measuring is aimed at determining the performance based on some kind of criterion, whereas benchmarking is the activity to compare the resulting scores with some kind of norm. Norms are often derived based on measuring results of other organizations. This type of comparison requires that similar measures can be used among organizations. The results of M&B activities should result in organizational improvement and stimulate organizational learning. Yet many M&B methods do not provide any learning and there is much criticism (see for example Bannister, 2007; Janssen, Rotthier, & Snijkers, 2004; Peters, Janssen, & Engers, 2004). Despite the criticisms, limited attention has been given to develop foundations and guidance for developing improved methods.

Stages-of-growth models are often used to represent the current status e-government (Peters et al., 2004). Often stages are modeled with sequential steps showing the growth, whereas many models are incongruent with each other (Lee, 2010; Siau & Long, 2005). The position within a certain stage is ideally determined using measures and the benchmarking norms are determined by measuring other organizations. Whereas a sound measurement model might be thought as the basis of stage models, often this is not the case. Stage-of-growth models are often based on intuitive appealing models without providing any guidance to determine in which stage an organization is. Often the focus of measurement is on a generic level at the expense of detailed insights (Bannister, 2007; Kunstelj & Vintar, 2004). This difficulty might result in the adverse effects that benchmarks might have limited practical meaning, but might have a huge impact on political decision-making (Bannister, 2007). The measurement is mainly based on the outcome of the past events, is often rather abstract and hardly takes the size, scope, and complexity of the government organizations into account (Bannister, 2007; Gupta & Jana, 2003). This makes it hard to understand the position of an organization and to identify opportunities for improvement. In conclusion, M&B and stages-of-growth models provide hardly any support for organizational development. They can only be used as a start for organizations by giving a global idea about the current position, but leave a void which actually can and should be done and what specific areas are those that need to be improved.

The *goal* of this paper is to develop a M&B method that provides organizations guidance in their development to a higher maturity.

* Corresponding author.
  E-mail addresses: d.maheshwari@tudelft.nl (D. Maheshwari),
m.f.w.h.a.janssen@tudelft.nl (M. Janssen).

Current measurement models provide either high-level views or are focused on the particular aspects and provide limited support for organizational development. This paper extends the framework of the paper of Maheshwari, Janssen, and Veenstra (2011) published at the ICEGOV2011 conference by including analyses of practical and theoretical challenges and introducing seven principles for architecting stages-of-growth and M&B models. In Section 2 literature is reviewed and existing M&B models are evaluated. In Section 3 the research method is presented. The multi-level measurement framework and constructs are presented in Section 4. In Section 5, the framework is illustrated in a case study at the Inland Revenue (IR) Karachi, Pakistan. Finally, in Section 6 we propose foundations for M&B and draw conclusions and recommendations.

2. Challenges in measuring and benchmarking

In literature, all kinds of criticism on M&B and stage models can be found, but there is no systematic overview available. Although there are various overviews of stage models (Kunstelj & Vintar, 2004; Lee, 2010; Siau & Long, 2005), there is no analysis of the difference in approaches they take and how they are dealing with shortcomings. We create an overview of criticism and how these are addressed in existing models by reviewing the literature and providing an overview of M&B challenges, and then use these challenges to evaluate existing models. This analysis will show that a lot of models copy each other's methods, but hardly address the criticism and challenges. They merely apply the same concepts to other areas without developing new insights.

2.1. Criticism on M&B and stage models

There are various comparisons of e-government measurement models and methods in the literature. These comparisons include the e-government models developed in practice as well as in research by academics, practitioners, governments, and international organizations. Ojo, Janowski, and Estevez (2005) compared three different surveys, those by the United Nations (UN), Accenture and Brown University to distil out a 'core' set of indicators. Janssen et al. (2004) identified 18 benchmarks in four areas e.g. supply studies, demand studies, information society studies and e-government indicator studies. Kunstelj and Vintar (2004) compared a large number of e-government measurement models used in different countries to show the diversity of these approaches. They analyzed 41 different measurement models focusing on 4 broad e-government aspects i.e. e-readiness, front-office, back-office, and effects and impacts. These broad aspects are further classified into sub-categories i.e. the aspect front-office is classified into demand and supply and the aspect e-readiness into government and citizens and businesses. Each model focuses on certain different aspects and their sub-categories. They found that the central focus of these approaches is the technical aspects with partial annexation of organizational and social aspects which are often much harder to measure as they concern 'soft' aspects. In a similar vein there exist a variety of interoperability measurement models with pre-dominant bias towards technical aspects while hardly including any organizational or social aspects (Group, 1998; Turnitsa, 2005; van der Veer & Wiles, 2008).

Bannister (2007) found that majority of measurement challenges are due to the lack of theoretical foundation, absence of methodological approach, adverse effects of high costs, absurd benchmarking based on end results with no process evaluation, and unstructured implementation methods. He described that the inadequate design and development of M&B methods pledge falsifying the government policies and have adverse cost effects. Coursey and Norris (2008) criticize the majority of e-government models for having weak empirical foundations. They argue that the models clearly lack methodological guidelines and remain intuitive, presumptive and speculative. The

origin and development of these measurement models come primarily from practice, and they are derived without employing structured and systematic research methods. These models are incongruent with each other as these are based on different perspectives and use somewhat different metaphors (Lee, 2010). Based on literature review of these measurement approaches and their critiques, we classified the criticism using categories of challenges in Table 1. We make a distinction between theoretical and practical challenges, as these elements should be solved differently. Theoretical challenges should be tackled by developing new knowledge and insight, whereas practical challenges need to be dealt within the actual M&B activities.

Although M&B is given a lot of attention in e-government, there are other areas in which models can be found including quality models (EFQM, 2003; Mayor, 2003; Nabitz & Klazinga, 1999), enterprise architecture maturity models (Schekkerman, 2006), capability maturity model (Paulk, Curtis, Chrissis, & Weber, 1993, 2002), and interoperability maturity models (Clark & Jones, 1999; van der Veer & Wiles, 2008). Although the design and development of these models vary from each other, it is striking that they face relatively similar measurement problems.

2.2. Evaluating e-government stages-of-growth models for M&B challenges

Despite the criticism, stages-of-growth models are still very popular, accepted in various areas and are adapted and used by a wide variety of stakeholders. The literature about criticism is often rather general and does not refer to specific models. Hence, we evaluated the stage models based on the theoretical and practical challenges as shown in Table 1. This should advance our insight if and how stage models have been adapted to the criticism. The selection of stage models is based on a number of citations, diversity and application. Table 2 shows the evaluation selected stage-of-growth models developed between 2000 and 2012. The M (main) indicates that one element is captured and discussed and S (supplemental) indicates that it was mentioned but not given any in-depth analysis or explanation. The characteristic measurement focus in Table 2 shows that the majority of models predominantly focus on the front-office with limited attention towards the back-office and incorporating social aspects. The overview shows that West (2004) and Andersen and Henriksen (2006) models focus mainly on front-office aspects, whereas the Wescott (2001) and Klievink and Janssen (2009) models on the back-office aspects. Although the latter focus on the organizational capabilities, the stages dominantly reflect back-office characteristics. The majority of the surveyed stage models have limited M&B and have a weak theoretical foundation. Most models that measure focus on technology, functionality and features that are visible, but do not include more social aspects like capacity, capabilities, readiness, commitment, user satisfaction and so on. The models merely measure what is there and do not look a step deeper in the organization.

The table also shows that almost all models do not have any empirical or theoretical underpinning. Notable exceptions are Ojo et al. (2011), Janssen and Veenstra (2005), Klievink and Janssen (2009) which use theoretical notions to derive the stages and Siau and Long (2005) and Lee (2010) who conducted a meta-synthesis of existing models. There is often no systematic approach to derive the models and the models seem to pop up at a certain moment in time. The comparison of the stage models reveals that the majority of stage models designed and developed at different time intervals share similar characteristics without any significant advancements being made by addressing the challenges and building knowledge. Finally, Table 2 shows that the stages-of-growth models hardly focus on the organizational level, but remain rather general. As such they provide little help for organizations to develop further. The models do not measure organizational details and do not capture

Table 1
Overview of theoretical and practical challenges.

No.	Challenges	Explanation	Authors
Theoretical challenges			
1	Lack of theoretical and empirical base	The existing stages-of-growth models are designed and developed in isolation by researchers, practitioners, and private, and public organizations. They do not include any research method to combine research with practice, lack methodological guidelines and remain intuitive, presumptive, and speculative.	Janssen and Veenstra (2005); Andersen and Henriksen (2006); Bannister (2007); Coursey and Norris (2008); Klievink and Janssen (2009); Maheshwari et al. (2011); Ojo et al. (2011)
2	Use of shallow and incomplete measures and limited understanding	A one-size fits all approach is often taken. Needs of one organization are often different (at least partially) from others. Existing e-government models focus on limited number of aspects that remain too vague, are ambiguous and do not provide in-depth understanding. The measurement models do not measure deep within the organizations, capture not all relevant data and provide limited understanding of what is really going on.	Janssen et al. (2004); Kunstelj and Vintar (2004); Maheshwari et al. (2011); Petrović et al. (2012)
3	Focus on a limited number of aspects	The majority of existing models focus on easy-to-measure aspects. Often these are hard or technical e-government aspects, but they fail to measure soft and socio-technical aspects that are often much harder to measure.	Andersen and Henriksen (2006); Bannister (2007); Maheshwari et al. (2011); Maheshwari and Janssen (2012); Grönlund (2011)
4	No improvement support	The focus is often on measuring and the resulting outcome, but the learning or feedback loop which start by M&B the situation is not included. Problem identification and understanding, diagnosis, identifying improvements, implementation and measuring the effects are neglected. All too often only numbers are compared, whereas the goal of benchmarking is to learn from each other.	Maheshwari et al. (2011); Petrović et al. (2012)
Practical challenges			
5	Utilization of M&B resources and cost	Limited resources and money limit the extent of measurement. Limited resources might result in outcomes that might not be valid or helpful.	Bannister (2007); Kromidha (2012)
6	Measurement of proxies resulting in ambiguous M&B interpretations	Measures often focus on available or easy to measure proxies. Instead of deriving measures from the objectives, measurement is driven by the available data. Operationalizing the measures require abstraction that might not prove to be valid.	Bannister (2007); Peters et al. (2004); Petrović et al. (2012); Kromidha (2012)
7	Ambiguous performance metrics	Measures are ill-defined resulting in misinterpretation, vagueness and oversimplifications. There is a mismatch between the original purpose and subsequent use of the results.	Janssen and Davidse (2004); Bannister (2007); Ojo et al. (2011)
8	Not able to deal with variety	Public organizations varying in size, scope, and complexity. Measures are high-level and do not capture the details and nuances between organizations. There is a reliance on a single measurement instrument (often internet browsing) that might not measure the variety of aspects.	Peters et al. (2004); Janssen (2010); Maheshwari et al. (2011)

aspects that require deep knowledge of organization. A possible explanation for this is that the M&B methodologies employ measurement techniques with focus on the easy to measure elements that can be observed from a distance. It is hard to say anything about a specific organization if there is no notion about the organization's objectives, its processes, systems, employees, client base and other aspects. Without measuring deep within organizations, it is only possible to remain at the generic level which does not help organizations to improve.

Overall most models have weak theoretical and empirical bases. Furthermore, often models are presented but these models are not tested or validated. Table 2 shows that current models hardly address the fundamental criticism on measurement and stage models and they merely copy each other's concepts without providing attention to the foundations of these models. Most of the criticisms on these models are well known, but there is hardly any theory development that helps to overcome these criticisms. Even meta-synthesizing approach taken by authors like Siau and Long (2005) and Lee (2010) does not contribute to the foundations. Although the resulting models may have some advantages over the previous models, it does not address the M&B challenges, and does not contribute to our understanding of M&B. Meta-synthesis results in merely yet another model and does not address the fundamental challenges or provide any concrete guidance for organizations. There are no rules of guidelines that help to develop better models. Despite the large amount of research in stages-of-growth and M&B there is limited contribution to the foundations. One reason for this can be found by looking at its origins. Measurement models to evaluate maturity or growth of

e-government are developed by individual researchers, practitioners, and public as well as private organizations. Often they come from different backgrounds and these models serve (slightly) different purposes. The design and development of these models are the outcome of different investigations and understandings of e-government, but not the result of thorough research. There is no theoretical guidance for developing these models.

3. Research method

Despite vast amount of criticism on the many models, there is little attempt to advance the knowledge about M&B and stage models. Although some models try to overcome the criticism (see for example Andersen & Henriksen, 2006; Klievink & Janssen, 2009; Ojo et al., 2011) most current measurement models only provide high level classification of stages, not always provide measurement support and have limited guidance for organization development. Furthermore they often lack sound theoretical and empirical underpinnings. Measurement is complex and requires in-depth understanding of the public organizations, their environment and developments.

The goal of this paper is to develop a comprehensive M&B method that provides organizations guidance for development and at the same time the paper should contribute to the foundations of developing M&B and stage models. Fig. 1 shows the research steps that were taken to develop a M&B method. The model comprises the 5 basic steps respectively. The first step is the identification of challenges based on literature survey (Section 2). Next our M&B model was constructed by overcoming the challenges as identified in step 1. This resulted in our

Table 2
Measurement characteristics of e-government stages-of-growth models.

No.	Measurement models	Use of methodology		Benchmarking measures		Measurement aspects		Measurement level			Measurement focus	
		Empirical foundation	Theoretical foundation	Fixed measures	Diverse measures	Organizational aspects	Technical aspects	National	Domain	Case	Front-office	Back-office
1	Baum and Di Maio (2000)	S		M		S	M	M	S		M	
2	Deloitte Research (2001)	S		M		S	M	M	S		M	
3	Layne and Lee (2001)	S		M			M	M	S		M	S
4	Hiller and Bélanger (2001)	S		M		S	M	M	S		M	S
5	UN (2001)	M	S	M	S	S	M	M			M	
6	Wescott (2001)	S	S	M		S	M	M	M		S	M
7	Moon (2002)	S		M		S	M	M	M		M	S
8	Technology (2002)	S		M		S	M	M			M	
9	Rohleder and Jupp (2003)	S		M			M	M	M		M	
10	UN (2003)	M	S	M	S	S	M	M	S		M	S
11	West (2004)	S		M			M	M			M	
12	Siau and Long (2005)	S		M		S	M	M	S		M	
13	UN (2005)	S	S	M	S	S	M	M	S		M	S
14	Andersen and Henriksen (2006)	S	S	M			M	M			M	S
15	UN (2008)	M	S	M	S	S	M	M	S		M	S
16	Klievink and Janssen (2009)	S	M	M		M	M	M	M		M	M
17	UN (2010)	M	S	M	S	S	M	M	S		M	S
18	Kalampokis, Tambouris, and Tarabanis (2011)	S	S	M		M	M	S	S	S	S	M
19	UN (2012)	M	S	M	S	S	M	M	S		M	S

initial measurement model which was primarily based on existing literature. This model included the measurement framework and constructs. In the third step this model was further refined using experts. A directive qualitative research method was used to refine the initially proposed theory in practice by conducting 16 interviews, which were analyzed using content analyses. The experts' comments and suggestions resulted in a further refinement of the model. The refined measurement model is discussed in Section 4. In the fourth step the refined measurement framework and constructs were used and tested in practice. This was done by conducting a measurement and benchmarking case study which is described in Section 5. The actual M&B was done using semi-structured surveys with participative sessions for collecting data and by applying qualitative research method using summative analysis approach. This step includes multiple methods; both interviews and group sessions were conducted. In the final step the use of the M&B model in practice is evaluated and conclusions were drawn. This will be discussed in Section 6.

4. Towards a comprehensive measurement and benchmarking model

In this section, the concepts and details of our M&B models are presented. Due to the scope of this paper, we do not include details about the data analysis of the refinement process and present only the refined model. The M&B models consist of a multi-level measurement framework, a range of measurement construct, data collection based on multiple methods and a dashboard to enable continuous improvement. Section 4.1 describes the refined version of the multi-level measurement framework. Section 4.2 discusses the development of measurement constructs by combining research and practice. Section 4.3 explains how multiple methods and information sources can be used to avoid the bias and create a more comprehensive picture of the situation at hand. Section 4.4 proposes the M&B

dashboard with simplified interpretation support to help management better understand the M&B outcomes and take correct measures.

4.1. Multi-level measurement framework

There can be diverse stakeholders that want to benefit from the M&B results. For the UN index this could be countries, but often these are organizations that want to know how well they are doing and how they can improve their situation. Yet organizations have a variety of objectives, have different natures and want to know different types of things in the benchmarks. In addition, measures might change due to technical and organizational developments. To foster this diversity in organizations a multi-level measurement is introduced. The use of levels is common in systems theory and is a mean to deal with complexity and to provide various levels of abstraction and detail (Jackson, 2003). In Fig. 2 the levels are classified from *generic* (containing measures that are applicable for *all* organizations), *domain* (containing measures that are applicable for all organizations within a certain domain, like for instance social security) to *organizational* specific indicators. The latter are only relevant for the organization under study and allows for the creation of a customized measurement instrument if needed.

Generic, domain and organizational indicators can be clustered in certain areas like for example backend and frontend. By making use of areas, M&B can be done only within a chosen area or M&B can be kept deliberately broad and a wide range of areas can be included. By utilizing this type of framework, flexibility for organizations is created to choose the different types of indicators, which are truly relevant for their situation. One can opt for benchmarking only within a certain area or only on a generic level or organizations can decide to utilize more resources and measure all the way to the organizational level. The disadvantage is that not always all information coming from all domains and areas will be collected and consequently there will be
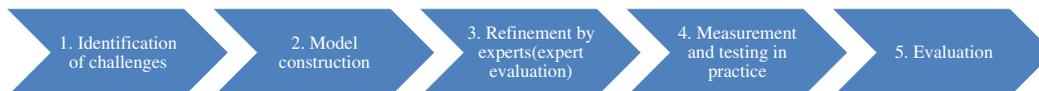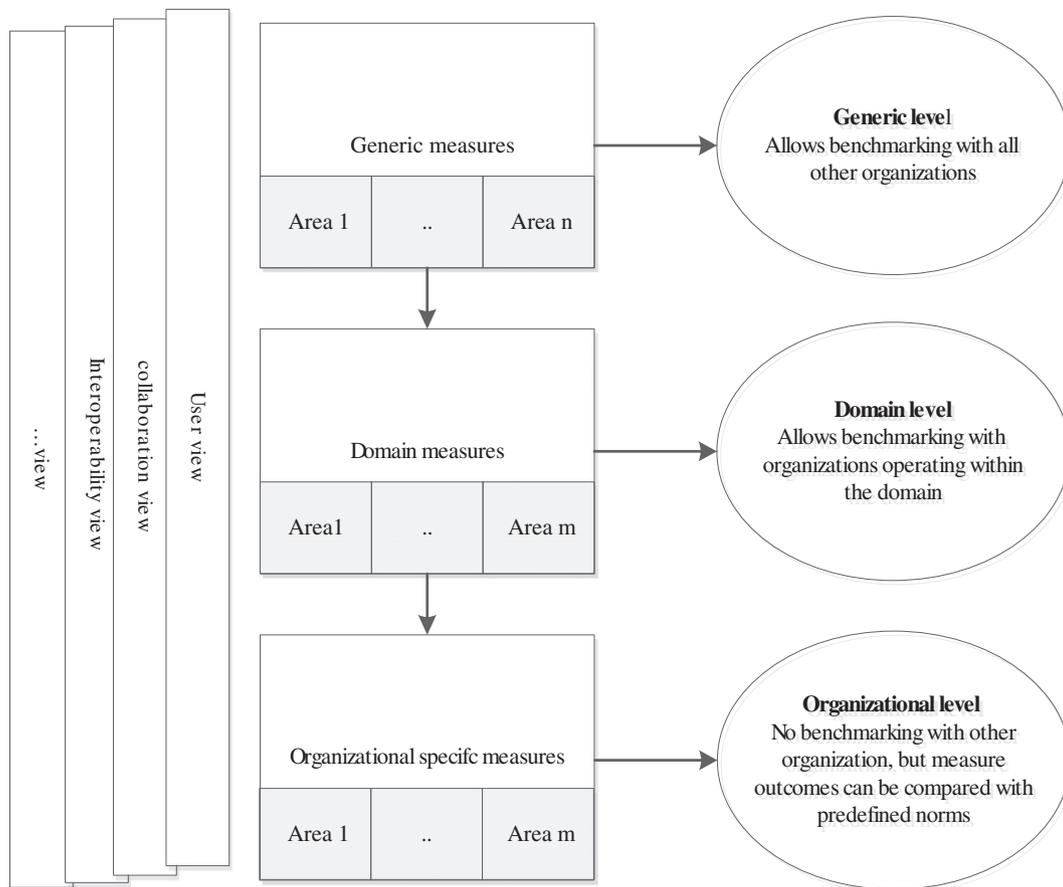

Fig. 1. Research steps.

Fig. 2. Multi-level measurement framework.

less data available that can serve as a benchmark. On the other hand, organizations might set their own norms and objectives and can use these as the norm for their M&B exercise. The advantage is that by focusing on levels and areas it is possible to zoom in upon the typical challenges an organization faces and avoid the waste or resources for measuring things that are not relevant at the end. The framework creates flexibility to focus on certain levels or areas and can function as a framework for selecting and narrowing down M&B activities. In this way, policy-makers might become more aware about what they really want to measure and benchmark.

The terms stages and levels are frequently used in the measurement models to classify different measurement aspects at different intervals in a systematic linear order. Maheshwari et al. (2011) found that though these stages or levels are classified systematically in linear fashion in existing measurement models, they are implemented on top of each other in practice. Instead of stages or growth levels, we prefer to use measurement constructs and variables to determine the level of sophistication. A stage or growth level only depicts the maturity in one area, whereas in other areas the maturity might differ. The use of stages can result in a narrow focus on achieving a certain stage, without looking at the rational and other areas that might be more important. Therefore, we prefer to drill down to areas and to utilize a variety of measure constructs and variables to understand the level an organization is operating at within a certain area. This can provide much better help for decision-making.

M&B methods often contain only a single view, whereas multiple views can provide much more insight. A *view* is "a representation of a whole system from the perspective of a related set of concerns" (Architecture_Working_Group, 2000 p. 2). Examples of a view are the user, security, interoperability or adaptability view. A viewpoint

establishes the rule conventions by which a view is created, visualized and analyzed (Architecture_Working_Group, 2000 p. 4). These views typically contain elements at various layers and are therefore added as a vertical column in the multi-level framework. Views are applicable for all areas, for instance security is relevant for both the front-office and back-office, but front-office requires different types of security than the back-office.

By utilizing areas and views, this measurement framework enables organizations to select specific relevant constructs and indicators from a variety of measurement methods developed for different areas. In this way, indicators that have been proven useful and valid in certain areas can be integrated in the method. Based on literature review we derived measures from various areas to arrive at a more comprehensive set of measures.

Having general and customized measures can also help to compare situations across countries. Though the number of domains within an organization may be similar among countries, the type of measurable information for these different domains can vary at large from each other. As such, generic, domain and organizational level measures should be included to allow measures that are generic and customized.

Historical benchmarking by comparison over time requires that metrics are time invariant, which is often not the case for all measures. Nevertheless, some measures might be time invariant. By employing areas and views certain measures can be kept deliberately constant over time to allow trend watching and observing growth over time, whereas other measures might be changed, removed or new measures might be added. In this way, changes at a particular level or in a particular domain or view will not affect the overall M&B procedure.

4.2. Measurement constructs and variables

In various areas, there exist already measurement constructs and variables that can be used for M&B purposes. Instead of developing our own constructs and variables, we opted for using measures originating from various areas and only developed our own if none were available. In this way indicators are used that are already proven in other domains, but still we were able to extend and customize these — if necessary. Our focus was on the back-office of public organizations. For our M&B method the constructs are derived from the different measurement methods as found in literature about stage models (Hiller & Bélanger, 2001; Lee, 2010; Reddick, 2004; Siau & Long, 2005), quality models (EFQM, 2003; Kaplan & Norton, 2001; Mayor, 2003; Nabitz & Klazinga, 1999), capability maturity model (Paulk et al., 2002), and enterprise architecture models (Schekkerman, 2006).

Stages-of-growth models dominated the e-government field during the last decade, whereas recently other measurement methods have been applied by public organizations to achieve the measurement objectives. A list of generic constructs by investigating and cross comparing the existing measurement approaches was developed. Identifying and comparing these measurement constructs at organizational level are difficult as the organizations vary at large depending on their scope, size, and complexity of day-to-day processes, procedures, and operations. Therefore, benchmarking at organizational level is complex and may not be cost effective, whereas domain and generic level benchmarking is relatively less difficult, and provides better chances to compare with the sister organizations.

In Table 3 an overview of the measurement construct is provided for measuring the back-office. In addition, the relationship between the derived constructs and the measurement approaches is shown. For example, the constructs two-way transaction and integration are derived from the stage models (Deloitte, 2001; Layne & Lee, 2001; Moon, 2002) and the constructs formal communication, collaboration, and syntax, semantic, and pragmatic are derived from the interoperability maturity models (Clark & Jones, 1999; Turnitsa, 2005). As some of the constructs overlap and can be identified in other measurement approaches, this classification should be merely viewed as an example. It does demonstrate that constructs are discussed in a sparsely scattered manner by different measurement methods and models.

The development of the constructs shown in Fig. 3 identifies the most relevant measures by combining research and practice. The initial list of constructs was later refined in our case study by conducting 16 interviews at Inland Revenue (IR) Karachi. IR Karachi is part of the IR Wing of Federal Board of Revenue (FBR) Pakistan and responsible for tax facilitation services in Karachi.

4.3. Use of multiple methods

Many measurement methods only use one type of data collection method, whereas in research triangulations are often emphasized to get a better and comprehensive picture from various fields and to increase the credibility and validity of the results. Often a single method is not able to capture all types of measures. By using multiple methods which confirm each other's findings there is much more proof that the answer is correct than using a single method.

Our multi-level measurement model allows public organizations to use multiple methods and information sources to avoid bias during data collection. In our method we used multiple information sources i.e. interviews, archived reports, news briefs, publications, and press releases to cross-check the data collected from surveys and participative sessions. Finally, the data was analyzed using the summative qualitative research method (Hsieh & Shannon, 2005).

4.4. Interpretation support: dashboards

M&B should not be a one-time activity resulting in some actions. Instead, it should be a continuous process of monitoring, benchmarking and improving. For this purpose, there is a need for a clear visualization of the various aspects that enables organizations to evaluate the effects of their actions, understand where they stand and select improvement activities. Visualization should be supporting interpretation and avoiding the preposterous consequences of misinterpretation. An index showing only M&B numbers (e.g. UN readiness index) does not provide enough insights and can easily be interpreted differently from its original meaning (Bannister, 2007). Even if the numbers are presented in different ways (Petrović, Bojković, Anić, & Petrović, 2012) it provides limited understanding. Therefore, there is a need to visualize the benchmarking scores of a variety of measures and aspects of the various areas and views.

Dashboards can be used to support interpretation of the outcomes and can be continually updated if new information is collected. Dashboards are aimed at showing the historical and current situation from a number of different views, to support its interpretation and to help to identify improvements. The term "dashboard" is commonly used in motorcycles, cars, and airplanes to show performance indicators i.e. speed, distance, and temperature. A popular definition of dashboard can be found in the book "Information Dashboard Design" by Stephen Few (2006) which definition was originally published in Intelligent Enterprise magazine article Dashboard Confusion. He defined a dashboard as "a visual display of the most important information needed to achieve one or more objectives, consolidated and arranged on a single screen so the information can be monitored at a glance" (Few, 2004, p-31). This definition confirms the need to focus on certain objectives and to visualize the outcomes in an easy-to-understand manner; however, the definition does not contain interpretation and improvement support. For continuous improvement, it is important that the cycle of measuring and improvement is followed up by another

Table 3
Back-office measurement constructs and their relationship with measurement approaches.

| No.	Measurement constructs (name)	Focus (levels)	Construct origin
1	Applications and services	Generic level	Stages-of-growth models
			Enterprise architecture
2	Two-way transaction	Generic level	Stages-of-growth models
3	Integration	Domain level	Stages-of-growth models
4	ICT infrastructure	Generic level	Stages-of-growth models
			Enterprise architecture
5	Strategy and policy	Domain level	Quality models
			Enterprise architecture
6	Service level agreement	Domain level	Quality models
			Enterprise architecture
7	Financial constraints	Organizational level	Quality models
8	Information privacy and security	Generic level	Quality models Interoperability models
9	Jurisdictional regulations	Domain level	Interoperability models Quality models
10	Syntax, semantic, and pragmatic	Domain level	Interoperability models
11	Formal communication	Organizational level	Interoperability models
12	Collaboration	Organizational level	Interoperability models
13	Business and IT alignment	Organizational level	Enterprise architecture
14	Back-office education	Organizational level	Enterprise architecture
15	Coherency	Domain level	Enterprise architecture
16	Environment and ethics	Organizational level	Enterprise architecture
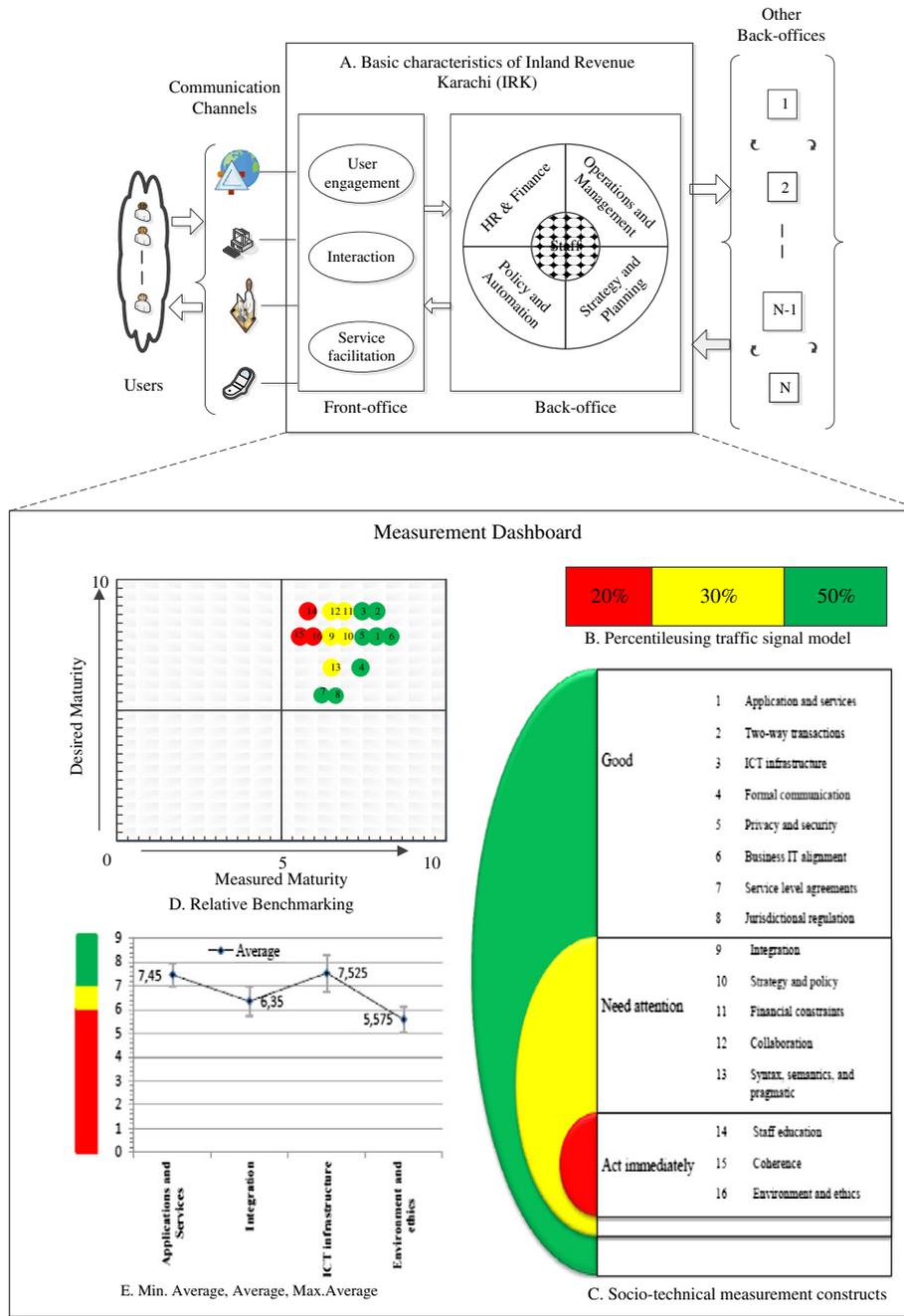
Fig. 3. M&B dashboard for the case study at IR Karachi.

measurement aimed at evaluating the impact of the improvement actions, which again should result in new improvement actions. Few's definition does not include the consequences of improvement actions and the effects. In our view, this is important as it facilitates understanding of actions and successful interventions can be used as suggestions for others. Therefore, we define a dashboard as "*a visualization of M&B information given a certain objective and providing interpretation and improvement support*".

Few (2006) also provides guidelines to avoid confusions and wrong utilization of dashboard, for instance necessary information is visualized in the boundaries of single screen, supplying sufficient contextual data, appropriate and eloquent presentation of data, and correct use of colors, pictograms, tables, charts, and graphs. DeBusk, Brown, and Killough (2003) use the term "dashboard measures" to

provide a quick view of organizational performance to the management. In a similar vein, our method uses M&B dashboards to graphically represent the measurement outcomes of the case study to simplify the interpretation ambiguities for policy makers, managers, administrators, decision makers, head of the departments (HODs), politicians and other stakeholders.

5. M&B method in practice

5.1. Background

This research used the proposed M&B model for assessment and benchmarking the IR Karachi. IR Karachi is part of IR Wing, which constitutes 90% of the total revenue collections by FBR Pakistan. It

has further three major divisions i.e. income tax, sales tax, and excise duty. Though the divisions are based on administrative operations and revenue collections, these taxes are somewhat similar in essence and highly interdependent on each other in practice. Therefore, a single administrative structure i.e. IR Wing was formed by FBR Pakistan to assimilate different taxes into an organized framework. IR Wing focuses on facilitating taxpayers to carry through their tax matters in an organized manner. This will also enable the easy access of relevant tax data from different domestic sections. In order to further facilitate citizens, businesses, and employees the IR Wing is divided into the regional tax offices (RTOs) and large taxpayer's unit (LTU) offices across the country. The case study at IR, Karachi comprises three RTOs and one LTU, which enabled benchmarking with each other. The measurement allowed the managers, and policy, and decision makers to determine the current status of the processes, procedures, and operations, whereas the benchmarking provided useful suggestions for improvements.

Measuring the status of maturity of the processes and procedures in IR Karachi is not easy, as there exists no measurement mechanism to determine the maturity of the on-going projects. In this case study, we used our measurement model comprising multi-level measurement framework, measurement constructs, a dashboard and utilizing multiple methods for measuring and benchmarking the maturity. The case study comprises semi-structured surveys and interactive participative sessions with the IR Karachi staff from RTOs and LTU.

5.2. Data collection

The surveys were conducted to measure the maturity of the proposed measurement constructs, whereas the participative sessions were arranged to evaluate the comprehensive measurement model. In total 40 surveys with the ICT experts, operational and technical staff, and managers were collected to measure the situation. The participants also explained that the organization is going through transformational changes due to organizational reforms. Small-scale participative sessions were arranged for evaluation of the measurement model. In total 2 participative sessions were arranged, each of them comprising 6 participants. The remaining people were encouraged to provide evaluation feedback via emails. In this way, multi-methods were employed consisting of 1) use of available data, 2) participative session, 3) interviews and 4) electronic feedback.

5.3. Dashboard to visualize the results

Fig. 3 shows a snapshot of the dashboard with the data of the case study surveys. The upper part of the figure gives an organizational impression of the IR, Karachi, whereas the lower dashboard provides an overview of the M&B outcomes. The M&B dashboard reflects 5 parts i.e.

1. basic characteristics of IR, Karachi,
2. percentile outcome of the measurement constructs using traffic light model,
3. individual aggregated measurement outcome of socio-technical measures,
4. relative benchmarking, and
5. an example graph depicting detailed descriptive analysis of measurement outcomes of the few measurement constructs.

5.3.1. Basic characteristics of IR Karachi

This part of the M&B dashboard represents the fundamental building blocks of the IR Karachi comprising the front-office, back-office, and links with the back-offices of other organizations, communication channels, and users. Our focus in this case study remains on the essential M&B aspects that reflect the overall maturity of the organization. This part of the diagram gives an overarching overview of the M&B scope and complexity in selecting the relevant constructs.

5.3.2. Percentile using traffic signal model

This visual representation using traffic model is utilized to simplify the M&B complexity. Instead of number, colors are assigned to the measurement ranges on the yardstick scale. The measurement outcomes are then aggregated and assigned to these ranges. As a result independent percentages of each aggregated range were determined. This combination of traffic model and aggregated measurement outcomes provides instant knowledge about the critical aspects of the IR Karachi. Thus, allowing management to quickly analyze the overarching current-status of the organizations.

5.3.3. Socio-technical measurement constructs

After the quick glance on the percentile reflecting the measurement outcomes, the managers, decision, and policy makers can switch to the next pictogram providing details about each measurement constructs. As the pictograms in the M&B dashboard are connected to each other by traffic model, the observers can quickly identify the individual constructs, their type (socio-technical), and current status. The list of constructs in this pictogram can be orally analyzed to identify the critical measures. Though the pictogram helps with simplified interpretation of the measurement outcome, the type of construct (organizational or technical) may require even deeper insights about the benchmarking data before taking any actions.

5.3.4. Relative benchmarking

The two dimensional pictogram in M&B dashboard shows the benchmarking outcomes, where measured maturity (horizontal axis) is plotted against desired maturity (vertical axis). Each participant was asked to reflect on the desired maturity of the measurement constructs on a 0-to-10 scale from low to high maturity. An average value for desired maturity was calculated for each measure and plotted against the measurement maturity by drawing intersecting horizontal and vertical lines. Each point of intersection representing the relative benchmarking outcome for a specific construct was assigned the concerned sequence number and color as shown in the previous pictogram. Though it looks like a simple benchmarking analysis, it can provide great deal of relevant details. For example the measured maturity levels for the constructs service level agreement and jurisdictional regulations were relatively low in the need attention $(6 < yellow \leq 7)$, range, but they were placed in good $(7 < green \leq 10)$ range due to the average low desired maturity levels in the need attention range.

5.3.5. Detailed example graph

Finally, the average graph in M&B dashboard is an example depicting detailed analysis of some measurement constructs showing average and standard deviation. The pictorial illustration of standard deviation in the graph represents the level of consensus and disagreements among the participants. This level of details can be very crucial to analyze the M&B variations in the participants' behavior and perception about measurement constructs, it depends on the scope and nature of research whether to include such details in the M&B dashboard.

5.4. Evaluation

The evaluation of the M&B exercise shows a number of benefits of the model. A main benefit is the use of multiple measurement levels containing generic, domain specific and case specific indicators. This enhances the benchmarking ability and at the same time allows for customization. All too often, the focus of measurement is at a generic level at the expense of detailed insights. Our multi-level measurement instrument can bridge this gap. Organizations are heterogeneous by nature as they operate within varying context and have different goals, staff, systems and processes. Benchmarking organizations based on a

single score has often limited use, whereas benchmarking on similar aspects can provide much more insight. The purpose of benchmarking is to improve and not to stare at the figures. The use of various measurement areas, including changes and transformation, usage, customer satisfaction, relation with other organizations, was considered as a huge help. This allows understanding of which areas were doing well and which not. This helps to understand organizations better and to concentrate on improvement activities.

The suggestions for improvement were viewed as essential, as this was viewed as the sole purpose of the B&M exercise. All agreed that dashboards are a valuable addition to the measurement and stage models. Especially the functionality to zoom in on details was appreciated as it allows having a benchmark at a glance, but still are able to analyze the reasons for the scores. Benchmarking should not only facilitate comparison, but should also enable learning from each other about what works and what does not work. In this way, a closed learning and feedback loop is created.

The use of multiple measurement methods was hardly mentioned as a benefit, as the use of multiple methods was hardly visible. The participants guessed that this would result in better measurement, but were largely unaware of this. They did mention that the use of multiple methods has the disadvantages that it consumes more resources and takes a longer time to complete.

The use of the method proved to be useful and beneficial for the IR Karachi. The measurement of the relevant socio-technical aspects deep inside the organization and visualization of the benchmarking data using M&B dashboard provided quick suggestions for necessary changes and useful insights for further improvement. Our case study shows that the process of measuring deep inside organizations might be more important than the actual outcomes. The measurement and benchmark is only a single activity that is part of the organization's continuous improvement activity. As such, it should not be viewed as an activity that is only done once, but it should be conducted annually to ensure that an improvement cycle is created. M&B should be followed up by improvement actions that in turn need to be evaluated. This process can be supported by dashboard, which can be used to evaluate the impact of actions and be updated when new measurement data becomes available. Only those areas with new data can be updated. Not all areas are addressed at the same time, as organization resources are often limited.

6. Conclusions

The research in this paper aimed at developing a M&B method that provides organizations guidance in their development towards higher maturity. M&B should not be viewed as an end, instead it should be considered as a means to improve organizations. The literature survey shows that many M&B and stage models copy each other's measurement methods, but do not address the limitations and challenges of M&B methods; they merely apply the same concepts to other areas. This might be a quick win, but does not advance our knowledge in the field of M&B and often results in yet another

Table 4
Foundational principles and their benefits and implications.

No	Foundational principles	Challenge addressed	Benefits and implications
1.	Taking into account organizational goals and public values	6. Measurement of proxies, resulting in ambiguous interpretation	– Focus on measuring what (what should be accomplished, instead of indiscriminately assuming that all organizations are similar)
2.	Distinguish between multiple measurement levels	2. Shallow and incomplete measurement and limited understanding 6. Measurement of proxies, resulting in ambiguous interpretation 7. Ambiguous performance indicators 5. Utilization of M&B resources and cost	– Accommodates benchmarking on generic indicators, while providing deep insight within domains – Allows public organizations to measure precise relevant information at the organizational level
3.	Use multiple measurement methods	2. Shallow and incomplete measurement and limited understanding 3. Focus on limited number of aspects 6. Measurement of proxies, resulting in ambiguous interpretation	– Combining various views resulting in better insight and recommendations – Avoiding bias and wring interpretations – Cost effective use of existing information sources for data collection – Benchmarking on broad range measurement aspects
4.	Define measures within various areas and views	2. Shallow and incomplete measurement and limited understanding 3. Focus on limited number of aspects 6. Measurement of proxies, resulting in ambiguous interpretation 7. Ambiguous performance metrics 8. Not able to deal with variety in organizations, processes, operations, cultures etc.	– Focusing on the different areas to gain deep insights – Addressing a broad range of measurement issues – Incorporating socio-technical organizational aspects in the measurement process – Allowing benchmarking of similar areas of different organizations
5.	Connect measures and stages-of-growth models (based on discontinuity principle)	1. Lack of theoretical and empirical base	– Use of concept of discontinuity (Cook, 1996) – Allows measuring the maturity of on-going operations – Provides clear measurement guidelines – Combines research with practice
6.	Benchmarks should provide suggestions for improvement	4. No improvement support 5. Utilization of M&B resources and costs	– Accommodates the future predictions and transformational changes in measurement process – Avoids adverse long-term M&B costs by updating specific measurement information – Allows implementation of the measurement constructs on top of each other
7.	Use dashboards to enhance communication and interpretation	1. Lack of theoretical and empirical base 4. No improvement support 7. Ambiguous performance metrics 8. Not able to deal with variety in organizations, processes, operations, cultures etc.	– Customization to the specific situation – Various measures are combined or integrated to support different views – Support interpretation and avoiding drawing wrong conclusions – Providing insight at a glance – Drawing attention to certain aspects – Measures and improvements can be integrated – Usable by non-experts – Incorporation of history to evaluate the impact of improvements

model. Derived from the criticisms as found in the literature eight theoretical and practical challenges of M&B and stage models were identified. The four theoretical challenges include 1) a lack of theoretical and empirical bases, 2) use of shallow and incomplete measurement and limited understanding, 3) focus on limited number of aspects, and 4) no improvement support. The four practical challenges include 5) utilization of M&B resources and costs, 6) measurement of proxies, resulting in ambiguous interpretation, 7) ambiguous performance metrics, and 8) not being able to deal with differences in organizations, processes, operations, cultures etc.

To overcome these challenges we developed a measurement approach consisting of a multi-level measurement approach, utilizing measurement constructs and variables derived from literature and visualizing the results using dashboards. Data collection is based on multiple data collection method instead of relying on a single method. This model uses areas and views to gain a more in-depth overview. The use of levels, areas and views in combination with multiple measurement methods helps to measure deep inside the organizations. Instead of having a single outcome, it is possible to see the performance per area or per view. In some areas or by taking a certain view, an organization might have a low score, whereas in other areas or views it might perform well. Organizations can benefit from the M&B method not only by knowing their position on the benchmark in the various areas they are involved in, but also by having specific guidance for improvement they can learn from the actions and interventions taken by others and the impact of these interventions. Benchmarking is more than comparing with a certain number. Benchmarking is about identifying and sharing practices with each other.

Developing measurement and stage models might be simple as demonstrated by the huge amount of models found, however, developing empirical and theoretical sound models is challenging. Based on the evaluation in the case study we abstracted seven foundational principles underlying our M&B method. Table 4 shows which principles are used to address which challenge and the resulting benefits of its use. Each principle contributes to the B&M and has different implications. These foundational principles help to derive stage models, measurement, and benchmarking models having a better foundation. In further research, these principles can be used to derive new or improved stage and M&B models and the principles can be further refined and extended. In particular, dashboards are hardly investigated by researchers and deserve ample research attention. For example, it is still unclear what types of dashboards are most suitable, which views should be included and how different stakeholders can make use of dashboards. In addition, the principles for visualizations of the M&B can be further refined and extended.

The use of the foundations might require the making of trade-offs to focus on certain aspects. For example, the use of multiple methods can increase the validity, but requires more resources than a single method. Policy-makers should make the trade-off between using more resources, overcoming the disadvantages and gaining the advantages. The framework in this paper helps to make such decisions, however, it does not offer support for making these decisions. In further research decision support can be created to use the foundations and make practical trade-offs. M&B is a complex phenomenon, which encompasses a broad range of theoretical and practical challenges. Using the principles can help to have better theoretical foundations, which avoids the development of yet another stage or measurement model, but the principles should not be viewed as an end.

Acknowledgments

References

Andersen, K. V., & Henriksen, H. Z. (2006). E-government maturity models: Extension of the Layne and Lee model. *Government Information Quarterly*, 23(2), 236–248.

Architecture_Working_Group (2000). *IEEE Std 1471-2000 Recommended practice for architectural description of software-intensive systems.* Washington, DC IEEE Standard 1471-2000.

Bannister, F. (2007). The curse of the benchmark: An assessment of the validity and value of e-government comparisons. *International Review of Administrative Sciences*, 73(2), 171–188.

Baum, C., & Di Maio, A. (2000). *Gartner's four phases of e-government model.* Stamford, CT Gartner Group (21, 12-6113).

Clark, T., & Jones, R. (1999). *Organisational interoperability maturity model for C2.*

Cook, M. A. (1996). *Building enterprise information architectures. Reengineering information systems.* New Jersey, USA Prentice Hall.

Coursey, D., & Norris, D. F. (2008). Models of e-government: Are they correct? An empirical assessment. *Public Administration Review*, 68(3), 523–536.

DeBusk, G. K., Brown, R. M., & Killough, L. N. (2003). Components and relative weights in utilization of dashboard measurement systems like the Balanced Scorecard. *The British Accounting Review*, 35(3), 215–231.

Deloitte, T. (2001). The citizen as customer. *CMA Management*, 74(10), 58.

EFQM (2003). *EFQM model for business excellence.* Brussels European Foundation for Quality Management 2003.

Few, S. (2004). Dashboard confusion. *Intelligent Enterprise*, 7(4), 14–15.

Few, S. (2006). *Information dashboard design: The effective visual communication of data.* O'Reilly.

Grönlund, Å. (2011). Connecting eGovernment to real government—The failure of the UN eParticipation index. *Electronic Government*, 26–37.

Group, C. I. A. W. (1998). *Levels of information systems interoperability (LISI).*

Gupta, M. P., & Jana, D. (2003). E-government evaluation: A framework and case study. *Government Information Quarterly*, 20(4), 365–387.

Hiller, J. S., & Bélanger, F. (2001). Privacy strategies for electronic government. In M. A. Abramson, & G. E. Means (Eds.), *E-government 2001* (pp. 162–198). Lanham, Maryland Rowman & Littlefield Publishers.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288.

Jackson, M. C. (2003). *Systems thinking: Creative holism for managers.* John Wiley & Sons, Inc.

Janssen, M. (2010). Measuring and benchmarking the back-end of e-government: A participative self-assessment approach. *Electronic Government*, 156–167.

Janssen, M., & Davidse, A. (2004). Evaluation of a performance-based accountability system. *Proceedings of the 4th European Conference on e-government* (pp. 403–410). Dublin Trinity College.

Janssen, D., Rotthier, S., & Snijkers, K. (2004). If you measure it they will score: An assessment of international eGovernment benchmarking. *Information Polity: The International Journal of Government & Democracy in the Information Age*, 9(3,4), 121–130.

Janssen, M., & Veenstra, A. F. v (2005). Stages of growth in e-government: An architectural approach. *Electronic Journal of eGovernment*, 3(4), 193–200.

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2011). Open government data: A stage model. *Electronic Government*, 235–246.

Kaplan, R., & Norton, D. (2001). Transforming the balanced scorecard from performance measurement to strategic management: Part I. *Accounting Horizons*, 15(1).

Klievink, B., & Janssen, M. (2009). Realizing joined-up government. Dynamic capabilities and stage models for transformation. *Government Information Quarterly*, 26(2), 275–284.

Kromidha, E. (2012, October). Strategic e-government development and the role of benchmarking. *Government Information Quarterly*, 29(4), 573–581.

Kunstelj, M., & Vintar, M. (2004). Evaluating the progress of e-government. *Information Polity*, 9(3–4), 131–148.

Layne, K., & Lee, J. (2001). Developing fully functional e-government: A four stage model. *Government Information Quarterly*, 18(2), 122–136.

Lee, J. (2010). 10 year retrospect on stage models of e-government: A qualitative meta-synthesis. *Government Information Quarterly*, 27(3), 220–230.

Maheshwari, D., & Janssen, M. (2012). Measuring organizational interoperability in practice: The case study of Population Welfare Department of Government of Sindh, Pakistan. *ICEGOV 2012, October 22–25, 2012, Albany NY, United States.*

Maheshwari, D., Janssen, M., & Veenstra, A. F. v (2011). A multi-level framework for measuring and benchmarking public service organizations: Connecting stages-of-growth models and enterprise architecture. *ICEGOV2011, September 26–28, 2011, Tallinn, Estonia.*

Mayor, T. (2003). *Six Sigma comes to IT: Targeting perfection.* CIO Magazine (available at: www.cio.com/archive (accessed 24 January 2004)).

Moon, M. J. (2002). The evolution of e-government among municipalities: Rhetoric or reality? *Public Administration Review*, 62(4), 424–433.

Nabitz, U., & Klazinga, N. (1999). EFQM approach and the Dutch Quality Award. *International Journal of Health Care Quality Assurance*, 12(2), 65–71.

Ojo, A., Janowski, T., & Estevez, E. (2005). Determining progress towards e-government: What are the core indicators? *Paper presented at the 5th European Conference on e-government, Antwerpen.*

Ojo, A., Janowski, T., & Estevez, E. (2011). Building theoretical foundations for electronic governance benchmarking. *Electronic Government*, 13–25.

Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. V. (1993). Capability maturity model, version 1.1. *Software, IEEE*, 10(4), 18–27.

Paulk, M., Curtis, B., Chrissis, M., & Weber, C. (2002). Capability maturity model, version 1.1. *Software, IEEE*, 10(4), 18–27.

Peters, R., Janssen, M., & Engers, T. M. v (2004). Measuring e-government impact: Existing practices and shortcomings. In M. Janssen, H. G. Sol, & R. W. Wagenaar (Eds.), *6th International Conference on Electronic Commerce*. Delft, The Netherlands ACM International Conference Proceedings Series.

Petrović, M., Bojković, N., Anić, I., & Petrović, D. (2012, October). Benchmarking the digital divide using a multi-level outranking framework: Evidence from EBRD countries of operation. *Government Information Quarterly*, *29*(4), 597–607.

Reddick, C. (2004). A two-stage model of e-government growth: Theories and empirical evidence for US cities. *Government Information Quarterly*, *21*(1), 51–64.

Rohleder, S., & Jupp, V. (2003). *E-government leadership: Engaging the customer.* New York Accenture.

Schekkerman, J. (2006). *Extended Enterprise Architecture Maturity Model (E2AMM).* (Version 2.0).

Siau, K., & Long, Y. (2005). Synthesizing e-government stage models — A meta-synthesis based on meta-ethnography approach. *Industrial Management & Data Systems*, *105*(3), 443–458.

Technology (2002). *The e-government handbook for developing countries.* Washington, DC Center for Democracy and Technology, World Bank ([Online] http://www.infodev.org)

Turnitsa, C. (2005). *Extending the levels of conceptual interoperability model.*

UN (2001). *Government Report (2001) "Benchmarking e-government: A global perspective-assessing the UN membestates".* UN Publication (http://www.upan1.org/egovernment2.asp)

UN (2003). *UN global e-government survey 2003.*

UN (2005). *Government readiness report 2005: From e-government to e-inclusion.* ONU. United Nations.

UN (2008). *Government survey 2008: From e-government to connected governance.* New York United Nations.

UN (2010). *United Nations e-government survey 2010.*

UN (2012). *United Nations e-government survey.*

van der Veer, H., & Wiles, A. (2008). *Achieving technical interoperability.* European Telecommunications Standards Institute.

Wescott, C. (2001). E-government in the Asia-Pacific region. *Asian Journal of Political Science*, *9*(2), 1–24.

West, D. M. (2004). E-government and the transformation of service delivery and citizen attitudes. *Public Administration Review*, *64*(1), 15–27.

Devender Maheshwari is a Ph.D. researcher at Delft University of Technology at the Faculty of Technology, Policy and Management. His research specializes in measurement and benchmarking of the public service organizations.

Dr. Marijn Janssen is the director of the interdisciplinary SEPAM Master Program, Master of Design and Compliance Management, and an associate professor in the field of ICT-systems service engineering and e-government within the Information and Communication Technology section of the Technology, Policy and Management Faculty of Delft University of Technology.