2014

Predictors of an Effective Performance Measurement System: Evidence from Municipal Governments in Turkey

Sedat Eliuz
University of Central Florida

PREDICTORS OF AN EFFECTIVE PERFORMANCE MEASUREMENT SYSTEM: EVIDENCE FROM MUNICIPAL GOVERNMENTS IN TURKEY

by

SEDAT ELIUZ
B.S. University of Ankara, 1999
M.A. University of Nottingham, 2002
Mag. rer. publ. German University of Administrative Sciences Speyer, 2007

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the College of Health and Public Affairs
at the University of Central Florida
Orlando, Florida

Fall Term
2014

Major Professor: Naim Kapucu

ABSTRACT

The aim of this study is to examine the predictors of effective performance measurement in the context of Turkish municipalities. In the study, mainly the theoretical guidance of context-design-performance model has been utilized to examine the contextual and design factors which have influenced the effective use of performance measurement systems in Turkish municipalities. The following research questions were examined in this study: To what extent do Turkish municipalities implement performance measurement systems effectively?, What are the predictors of effective performance measurement in Turkish municipalities?, and whether or to what extent do quality of performance measures, technical capacity of the municipality for performance measurement, organizational support, and external support for the use of performance measurement have influence on the effectiveness level of performance measurement systems in Turkish municipalities? In the study, the data were collected from Turkish municipalities by a self-administered online survey and were analyzed by using the structural equation modeling (SEM).

It is hypothesized in the study that external support and organizational support for the use performance measurement, and technical capacity for the performance measurement are associated with quality of performance measures and effectiveness of performance measurement systems in general. The results of the study supported the hypotheses of the study regarding the relationships among organizational support, technical capacity, quality of performance measures, and effectiveness of performance measurement. Although the results confirmed that external support has an indirect effect on effectiveness of performance measurement via technical capacity and quality of performance measures, the hypothesis regarding the direct effect of it on effectiveness of performance measurement was not supported. Moreover, the study found that support of employees and citizens for the use of

performance measurement in Turkish municipalities are relatively low, the municipalities have deficiencies both in the quantity and the quality of staff that are responsible for performance measurement activities, and the level of employee involvement in the development of performance measures is low.

I dedicate this study to my family. They were with me whenever I needed them.

ACKNOWLEDGMENTS

I would like to take the opportunity to thank those who helped me during this academic journey and made this dissertation possible. First and foremost, I would like to express my gratitude to the chair of my dissertation committee, Dr. Naim Kapucu, for his excellent guidance as well as continuous support and patience in regard to the overall process. I have really benefited from his inspiration, knowledge, and persistence, which, I think, will contribute to my academic and professional life significantly.

I would also like to thank the other dissertation committee members, namely, Dr. Lawrence L. Martin, Dr. XiaoHu Wang, and Dr. Qian Hu, for their insightful comments and guidance that were instrumental in completion of my dissertation. It was a pleasure and honor to work under the guidance of such distinguished professors and receive their invaluable support.

My gratitude is extended to all professors and administrative staff in the College of Health and Public Affairs at the University of Central Florida who have contributed me in achieving this doctoral degree. I specifically want to thank Dr. Thomas T. H. Wan, Dr. Youcheng Wang, Dr. Ning Jackie Zhang, and Dr. Thomas Bryer, who did not hesitate to share their knowledge with me and inspired my interest of learning.

Institutionally, I would like to thank the Turkish Government and Turkish Ministry of Interior for their financial support in this doctoral program. I also would like to thank my colleagues and friends from the Turkish Ministry of Interior for being with me whenever I needed them. They made my journey through the public affairs program a joyful and memorable experience.

Last but most important, I profoundly thank my dear wife, Betul, my kids, Mithat Eren and Emine Berra, my parents and parents-in-law, and our family friend, Hatice Baykal. They gave their full support during this study. Their loving patience enabled me to complete this study. I am grateful for the sacrifices each of them has made for me. Without the support of them, this dissertation would never have been completed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

AGFI	Adjusted Goodness of Fit Index
AKP	Justice and Development Party
CDP	Context, Design, and Performance Model
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CMIN	Model Chi-square
CMIN/df	Relative Chi-square
CSM	Covariance Structure Model
CSP	Context, Structure, and Performance model
DF	Degrees of Freedom
EU	European Union
GAO	The U.S. Government Accountability Office
GASB	Governmental Accounting Standards Board
GFI	Goodness of Fit Index
IFI	Incremental Fit Index
IMF	International Monetary Fund
IRB	Institutional Review Board
IT	Information Technology
OECD	Organization for Economic Cooperation and Development
NFI	Normed Fit Index
NNFI	Nonnormed Fit Index
NPM	New Public Management
PM	Performance Measurement
RFI	Relative Fit Index
RMSEA	Root Mean Square Error of Approximation
SEM	Structural Equation Modeling
SPA	Special Provincial Administration
SPO	Structure, Process, and Outcome Model
SRMR	Standardized Root Mean Square Residual
TGNA	Turkish Grand National Assembly
TLI	Tucker-Lewis Index
UCF	University of Central Florida

UK	United Kingdom
US	United States
WB	World Bank
$\chi 2$	Chi-Square
$\chi 2/df$	Chi-square / Degree of Freedom

CHAPTER 1: INTRODUCTION

1.1 Statement of the Problem

Measuring the performance of organizations has always been a concern of public administration since the beginning of the 20th century. However, it has not gained more than a medium level of attention until the last 2-3 decades (Streib & Poister, 1999). Especially beginning from the 1980s and 1990s there has been a transformation in the perspectives of public administrations in the world from a rule-based, process-oriented Weberian traditional public administration to a result-oriented market-based public management approach (Hughes, 2012; OECD, 1998), of which measuring the performance is an important part (Moynihan, 2006). As a result, the last three decades have experienced an increasing interest in using performance measurement systems in public administration throughout the world with the aim of improving accountability and performance (Abramson, Breul, & Kamensky, 2006; Berman & Wang, 2000; Hoontis & Kim, 2012).

Measuring the performance of public organizations was not considered as an important concern for Turkish public administration until the last decade. However, administrative reforms in Turkey between 2003 and 2005 have embraced the performance measurement as an important aspect and required the municipalities (among others) to measure the performance of municipal activities. According to Ates and Cetin (2004), the government agencies in Turkey have recently seen the introduction of performance related mechanisms as a way to overcome the problems in public administration and as a result there has been a significant increase in the number of laws and regulations, which contains the concepts such as performance measurement, strategic management, accountability, and transparency.

These concepts have also been included in both the Greater City Municipality Law (passed in 2004) and the Municipality Law (passed in 2005), which made preparing the strategic plans and measuring the performance of municipal activities and personnel obligatory for the municipalities. However, to what extent the Turkish municipalities have embraced these responsibilities and have been implementing these effectively is an issue of contention in the literature (Koseoglu, 2005; Sezen, 2011). Since the issue has not been examined empirically, there is a need for a study, which examines the current state of performance measurement activities in Turkish municipalities.

1.2 The Purpose of the Study

The aim of this study is to examine to what extent Turkish municipalities implement performance measurement systems effectively and which factors are important in the level of effectiveness of these systems. More specifically, the study aims to examine the role of stakeholder (both internal and external) support and technical capacity on the effective implementation of performance measurement in the context of Turkish municipal governments.

1.3 The Scope of the Study

In performance measurement systems, the performance of a unit is regularly and systematically collected, analyzed, and reported. This unit can be an individual, a group, a program, or an organization. This study is concerned only with the performance measurement in the organizational level. It explores the potential effects of four factors on the effectiveness of organizational performance measurement: external support (support of citizens and council members), organizational support (support of mayors, public managers, and employees), technical capacity of municipality, and quality of performance measures.

1.4 Research Questions

The following research questions are examined in this study: To what extent do Turkish municipalities implement performance measurement systems effectively?, What are the predictors of effective performance measurement in Turkish municipalities?, and whether or to what extent do quality of performance measures, technical capacity of the municipality for performance measurement, organizational support, and external support for the use of performance measurement have influence on the effectiveness level of performance measurement systems in Turkish municipalities?

1.5 Theoretical Perspectives

In the study, Context-Design-Performance (CDP) model, or context-design theory as called by Wang (2010), is used as the main theoretical guidance to examine the predictors of effective performance measurement in Turkish municipalities. This model is heavily influenced from contingency theory (Agiro, 2011; Burke & Litwin, 1992; Goltz, 2006; Marathe, 2006; Wan, 1995) and Donabedian's Structure-Process- Outcome (SPO) model (Agiro, 2011; Goltz, 2006; Wan, 2002). Contingency theory perspectives are useful in explaining the importance of external factors on organizational performance, whereas SPO model focuses more on internal factors. CDP model is a mixture of these perspectives and concerns both external and internal factors which influence the performance of an organization.

1.6 Significance of the Study

In the literature, measuring the performance of public organizations, which uses the authority to act on behalf of the public, is seen as an important way of holding them accountable (Behn, 2001; Hill & Lynn, 2009; Martin & Frahm, 2010; Moynihan, 2008). The

accountability dimension has been one of the most important motivations for the Organization for Economic Cooperation and Development (OECD) to promote performance measurement (Greiling, 2006). In addition to its contributions on improving accountability, performance information may also help strategic planning and budgeting practices (Hatry, 2006). Moreover, it has the potential of contributing the improvement of the services provided (Ammons & Rivenbark, 2008). However, the expected benefits of performance measurement can only be accrued with the existence of appropriate performance measures and their rigorous application. As succinctly put by Bouckaert (1993) "[i]t is necessary not just to focus on the measurement of performance but also on the performance of measurement" (p. 42).

This study builds on and contributes to earlier studies on performance measurement. Although earlier studies have examined the factors which influence the use and the effectiveness of performance measurement systems in the US public sector (Ammons & Rivenbark, 2008; Berman & Wang, 2000; de Lancer Julnes & Holzer, 2001; Mausolff & Spence, 2008; Melkers & Willoughby, 2005; Poister & Streib, 1999; Streib & Poister, 1999) and in other countries (Pollanen, 2005; Salazar & Martinez, 2013; Taylor, 2006, 2011; Yang & Hsieh, 2007), to my knowledge, the issue is not studied for the Turkish municipalities. Moreover, earlier studies did not examine the predictors of high quality performance measures. But this study also evaluates the quality of performance measures in addition to effectiveness of performance measurement systems.

Furthermore, as explained above, many researchers argue that Turkish municipalities do not implement performance measurement effectively (Koseoglu, 2005; Sezen, 2011). However, these arguments have not been examined empirically and therefore, there is a need for empirical studies in the literature. Since, as argued by Behn (2001) and Moynihan (2008),

performance measurement has the potential to improve the accountability and performance of the public agencies, examining the factors which affect the use of performance measurement in Turkish municipalities may also contribute to those ends. As the concept of "performance" is new in Turkish public administration, so are the studies conducted about the performance of Turkish public agencies. Therefore, this study contributes to filling a gap in the literature and may help policy-makers, elected officials, and managers reconsider and improve the effective use of performance measurement systems.

CHAPTER 2: LITERATURE REVIEW

2.1 Performance Measurement

Performance measurement has several definitions in the literature. The U.S. Government Accountability Office (GAO) (2011) defines performance measurement as "the ongoing monitoring and reporting of program accomplishments, particularly progress towards preestablished goals" (p. 2). In his definition, Hatry (2006) specifically mentions the measurement of results and efficiency of services and activities. Similarly, de Lancer Julnes and Holzer (2008) also point out the importance of regular collection of data on activities and accomplishments. Martin and Kettner (2010) emphasize both the "regular collection and reporting of information about the efficiency, quality, and effectiveness of programs" (p. 4). Based on these definitions, performance measurement can be defined for the purposes of this study as a management tool which encompasses the regular and systematical collection, analysis, and reporting of the performance of an organization.

Fryer, Antony, and Ogden (2009) point out the four aspects of performance measurement which are: "(1) deciding what to measure; (2) how [and when] to measure it; (3) interpreting the data; and (4) communicating the results" (p. 481). The first aspect is about developing relevant performance measures based on the objectives of an organization. The second aspect concerns how and when the performance measured. The third aspect is the transformation of performance data into useful information through the analysis of performance data. The last aspect is the communication of results to internal and external stakeholders in a way that enhances the effectiveness of performance measurement.

The notion of performance measurement, the idea that governments and other authoritative bodies can place hard data on the actions and services provided by them, has

been around since Frederick Taylor at the onset of the 20th century (da Cruz & Marques, 2014; Folz, Abdelrazek, and Chung, 2009; Holzer & Yang, 2004; Poister & Streib, 1999; Streib & Poister, 1999). Williams (2003) presents evidence that, in this period, New York Bureau of Municipal Research used performance measurement which contained many of the features of the current practice. But, performance measurement has not gained more than a medium level of attention in public administration until the last three decades. However, beginning from the 1980s and 1990s, performance measurement has become highly popular throughout the world. Behn (1995) counts the measurement of organizational performance as one of the three big questions of public management deserves researching. According to Dalehite (2008) "performance measurement continues to be one of the hottest topics in public administration today" (p. 891)

In the literature, the relatively extensive utilization of performance measurement has been linked to the widespread managerial reforms throughout the world (Fryer, Antony, & Ogden, 2009). Especially the last two decades of the 20th century "have seen a plethora of reinventing, rationalizing, reengineering and reforming initiatives designed to improve the organizational efficiency and effectiveness of the public service" (Kosecik, Kapucu, & Sezer, 2003, p. 105). Many developed countries including the US, the UK, other Western European administrations, Australia and New Zealand moved into an era of severe administrative reforms in this period (Kapucu & Palabiyik, 2008). The transformation of the public administrations of these countries from a Weberian traditional bureaucratic model to a market-based, efficiency-oriented, and effectiveness-oriented model has been explained by several scholars with different names, but as argued by Hughes (2012), literature has more or less settled on the name the New Public Management (NPM).

Performance measurement is an important aspect of these administrative reforms, which have taken place in the last three decades (Fryer et al., 2009; Hughes, 2012; Kapucu, 2010; Moynihan, 2008). Moynihan (2008) argues that the most frequent and widely adopted reforms of this period are tied to the concept of performance. By using the tenets of the NPM, Moynihan (2006) argues that:

> Moving from an administrative culture of compliance, error avoidance, and presumed inefficiency to a more efficient and effective public service requires multiple changes to existing formal systems. First of all, administrative goals should be specified through some sort of formal strategic planning. Short-term strategic goals are intended to be consistent with longer-term strategic plans for the organization. Goals are defined in measurable terms that compare ex-post performance to ex-ante targets. (p. 79)

Since the legitimacy of a government derives from the consent of governed, it is important that the government is accountable (Hughes, 2012). There are several researches about the importance of performance measurement in enhancing accountability and responsibility of government action (Zamesnik, 2012). For example, Behn (2001) mentions "accountability for performance" as a type of accountability in addition to "accountability for finances" and "accountability for fairness." Moreover, Martin and Frahm (2010) argue that financial accountability was the main concern in the discussions about public administration in the early times. However, the performance measurement movement united performance accountability and financial accountability and now "being accountable for the efficiency (outputs), quality and effectiveness (outcomes) is at the crux of administrative practice" (Martin & Frahm, 2010, p. 138). In a similar vein, Moynihan (2008) points out the importance of performance regarding accountability and argues that the performance information is important for accountability to the public, because it provides a transparent explanation of whether and how well the government is doing. Moreover, it contributes to

accountability to elected officials by reducing information asymmetry and helps them to exert oversight and improve their ability to direct public services (Moynihan, 2008).

A government cannot be effective, indeed cannot survive, without the trust (and implied financial contribution) of its citizens, and thus they must seek to improve citizen trust (Hoontis & Kim, 2012; Yang & Holzer, 2006). As a result, governments have attempted various reforms throughout the years to enhance their perceived image, utilizing a variety of initiatives. Whatever the reform attempt, the ultimate goal seems to be predicated on the idea that improved government performance will lead to a reversal of the decline of trust (Yang & Holzer, 2006).

Moreover, Hoontis and Kim (2012) argue that there is a widespread belief that performance measurement facilitates effective and efficient management. Performance measurement create information, which can be used by public managers to assess the level of organizational improvement, to diagnose the problems, and to make modifications in organizational strategy to respond to changing needs and priorities (Hoontis & Kim, 2012). Effective performance measurements must be implemented in order to gauge whether government programs are meeting their goals. Ho (2002) says that there is an increased interest by citizens to know where and how their tax dollars are being spent, if their requests are being heard, and if this translates into a benefit to the lives of an average citizen. In order to engage the citizen, a government must make performance measurement results available so that a transparent relationship is fostered, thereby reinforcing trust between the two (Ho, 2002).

In addition, performance information may also contribute to the formulation and justification budget requests; help the allocation of budgetary resources; trigger the detailed examinations of whether and why performance problems or successes exist in specific

departments; support strategic and other long-term planning efforts; and analyze options and establish priorities (Hatry, 2006; Holzer &Yang, 2004). Behn (2003) also mentions 8 purposes of public managers in using performance measurement: evaluation, control, budgeting, motivation, promotion, celebration, learning, and improvement. According to him, the other seven purposes are only means for achieving the real purpose, which is the improvement of the performance.

Folz, Abdelrazek, and Chung (2009) found that improving management decisions, supporting budget recommendations and decisions, and responding to citizen demands for greater accountability are the most important reasons why chief executives of the US local governments adopted performance measurement. Similarly, Poister and Streib (1999) reported that making better managerial decisions is the principal motivator for adopting performance measurement.

There are some in the literature who criticize performance measurement and the actual effective role it plays in the development of public policy (Moynihan 2006; 2008; Perrin, 1998; Streib & Poister 1999, Yang & Holzer 2006). As argued by Dalehite (2008), the extent the information produced by performance measurement systems is utilized by governments is dubious. The underlying theme here seems to suggest that though there are many jurisdictions who track performance measures, and who report their findings diligently, little substantive change results from the information gathered. For example, when it comes to allocation of resources, surveys of administrative officials have shown that the decisions tend to be political, rather than based on departmental performance (de Lancer Julnes & Holzer, 2001).

Several studies point out that if not implemented effectively, performance measurement may lead to goal displacement, meaning that performance indicators become as the goals (Moynihan, 2008; Perrin, 1998). Moreover, utilization of indicators which focuses

too narrowly on outcome and short-term financial situations may limit the effectiveness of performance measurement. In that sense, Perrin (1998) advocates the utilization of a broader approach, such as Kaplan and Norton's (1992) balanced scorecard method, that considers a wide range of indicators including difficult-to-measure factors such as focus on innovation and learning.

Moynihan (2008) argues that the creation, selection, interpretation, and presentation of performance information is not an automatic or objective process, but is influenced by the roles that actors in the political process occupy. There may be several pieces of information, but individuals can place more or less weight on this data. In government, there are often many performance measures that tell different stories about whether a program is successful; one piece of performance data is chosen over another depending on the perspective of the user (Moynihan, 2008; Radin, 2006).

In accordance with the increased interest regarding performance measurement and practice in public administration, the last decades have also seen a growing body of literature, which concerns the issues related to performance of public organizations (de Lancer Julnes & Holzer, 2001). However, as argued by Yang and Hsieh (2007), these studies are mostly descriptive and prescriptive. Most of the studies suggest and prescribe the important steps and processes in performance measurement; however, they did not empirically test the validity of these arguments (some exceptions include Berman & Wang, 2000; de Lancer Julnes & Holzer, 2001; Ho, 2006; Taylor, 2006; Yang & Hsieh, 2007).

de Lancer Julnes & Holzer (2001) see the utilization of performance measurement systems consist two stages: adoption of performance measures and implementation of performance measurement. Adoption stage refers to the development of a capacity of act, which in the context of performance measurement includes the processes of developing and

adopting performance measures. Implementation stage is the actual use of performance data, such as for strategic planning, funding, and decision making (de Lancer Julnes & Holzer, 2001). Several studies in the literature focused on one of these stages (Berman & Wang, 2000), whereas some others examined the both stages (de Lancer Julnes & Holzer, 2001; Yang & Hsieh, 2007).

de Lancer Julnes and Holzer (2001) evaluated not only the adoption phase of performance measurement but also the implementation phase. They examined importance of several factors, which they classified into two categories: rational/technocratic factors (such as information, resources, goal orientation, internal and external requirements, and goal orientation) and political/cultural factors (such as external and internal interest groups, risk taking culture, attitudes). They found that the rational factors are more important in the adoption level of performance measurement in US state and local governments, whereas the implementation phase is influenced heavily by the political factors. Yang and Hsieh (2007) also found that adoption and implementation are different constructs which have different determinants.

2.2 The Use of Performance Measurement Systems in the US Local Governments

Studies of performance measurement systems currently entrenched in city governments have revealed that many city officials find measuring performance a useful and worthwhile exercise (Wang, 2002). While it may not inform public policy or result in an allocation of resources, in a passive sense, reporting performance measures allows citizens to become involved, and gives managers the ability to point to past accomplishments (Wang, 2002; Yang & Wu, 2013). Wang (2002) found that performance measures help managers and other decision makers to identify service goals, strategies, and expectations. Additionally, in

theory, a comprehensive, refined performance management system should help policy makers make decisions regarding budgeting and programming (Zamesnik, 2012).

Similarly, Ho (2006) examined the perceptions of Midwestern mayors about the usefulness of performance measurement systems and found that it is useful if the performance data are not only used for reporting but also included in strategic planning, goal setting, and communication between city officials and organizational actors. Moreover, he found that involvement of major stakeholders in the process of developing performance measures increases the perceived usefulness of the tool. The findings of a national survey conducted by Folz et al. (2009) also supported the usefulness of performance measurement. The scholars found that most chief executives in medium sized U.S. cities (with populations between 25,000 and 250,000) thought that performance measurement system they used met or exceeded their expectations.

The use of performance measurement systems have been proliferated in the US local governments in the last two decades and most local governments use some kind of performance measurement (Hoontis and Kim, 2012). In their study, in which they collected data from city and county administrators from 47 counties and 168 cities, Melkers and Willoughby (2005) found that the use of performance measurement by the US local governments is pervasive. Almost half of the respondents reported that all of their departments use performance measurement. Moreover, another 20% reported that at least half of their departments use it (Melkers & Willoughby, 2005). The pervasive use of performance measurement systems in medium-size cities was also supported by the findings of Folz et al. (2009). The use of performance measurement is related to the population of local government as well (Salazar & Martinez, 2013). Several scholars found that larger cities are more likely

to adopt performance measurement than the smaller cities (Folz et al., 2009; Poister & Streib, 1999; Salazar & Martinez, 2013).

Another important factor which influences the level of the utilization of performance measurement systems is the form of government. Poister and Streib (1999) found that local governments with a council–manager form of government use performance measurement systems more than those with a mayor–council form. In a similar vein, Folz et al. (2009) found that cities served by a professional top manager use performance measurement more frequently (70% to 50%) than cities led by an elected official.

The local governments are using several performance measurement systems such as balanced scorecards, management dashboards, and operations-assessment tools. Balanced scorecard method, which is originally developed for the private sector by Kaplan and Norton (1992), is concerned simultaneously with several aspects of the management such as financial, customer, internal business, and innovation and learning perspectives. These perspectives provide a comprehensive view of the performance of an organization (Edwards & Thomas, 2005). In general, "[b]alanced scorecards, like the one used in Charlotte, tend to set strategic direction by tying together a loose set of management goals and philosophies" (Edwards & Thomas, 2005, p. 375).

Some other local governments have been using management dashboards, which "translate that strategic direction into a set of specific strategic outcomes that are tracked and monitored" (Edwards & Thomas, 2005, p. 375). One of the known examples of this method is Atlanta dashboard which was introduced in 2002. Different from balanced scorecard, Atlanta dashboard concentrated only in financial and customer satisfaction perspectives. The dashboard focuses on outcomes (such as reducing the crime, and reducing fire loss), rather

than inputs, and outputs and leaves latitude to departments in deciding how to achieve those targets (Edwards & Thomas, 2005).

Another method used in the US local governments to measure their performance is the utilization of operations-assessment tools. One prominent example of this is the CitiStat, which is used in Baltimore, Maryland. These operation-management tools "make sure that day- to-day operations are functioning properly and are focused on achieving those strategic outcomes" (Edwards & Thomas, 2005, p. 375). By using graphic mapping tools to track activities, CitiStat presents managers almost real-time data about the departmental activities (Edwards & Thomas, 2005).

Other than these methods which measure the performance within one municipality, some municipalities use benchmarking to create their common performance measurement system. Instead of a separate language of performance measurement, these municipalities create a common language in which common services and performance measures are utilized (Boyer & Martin, 2012). In their study, in which they examined the largest intrastate performance measurement consortium in the United States, the Florida Benchmarking Consortium, Boyer and Martin (2012) argued that the data collected for the Consortium is useful to compare how same services are delivered by similar local governments. Such a joint action can contribute the local governments in indicating the possible problem areas in the provision of their services.

2.3 Performance Measures

An important part of performance measurement systems is developing the measures which will be used to measure the performance. As argued by Wang (2010), developing the appropriate indicators is a prerequisite for the successful implementation of any performance

measurement system. However, as argued by Fryer at al. (2009), developing measures, their quality, and their reporting are still important problem areas in performance measurement practices. Developing measures in the public sector is even more difficult because of the conceptual problems about defining good performance and the role of the public sector (Van de Walle, 2008). Yet the use and quality of performance measures are evolving from being primarily financial, to gradually other measures, such as quality (Fryer et al., 2009; Johnsen, 2005; Rejc, 2004).

In the literature, generally four types of performance measures (inputs, process, outputs, and outcomes) or their variants (such as efficiency and effectiveness measures) are mentioned (Kapucu, 2010; Wang, 2010). Input measures assess the level of resources used to produce goods. Process measures are about the activities or workloads in the production process. Output measures concentrate on the amount of good produced or service provided and mainly seeks the answers to the questions like "how many" or "how much" (Ammons, 2013). As written by Ridley and Simon (1938):

> We can measure the miles of beat patrolled, the number of criminals apprehended, the number of finger-prints taken. But units such as these, however useful they may be, are not entirely adequate for our purposes. They tell us how much work has been done; but they do not tell how well it was done, nor whether the particular work undertaken was appropriate to the desired end. A measurement of the result of an effort or performance indicates the effect of that effort or performance in accomplishing its objective. (as cited in Ammons, 2013, p. 2)

Since many public agencies provide only services, not tangible products, their service delivery process can be considered as the service output. Therefore, process and output measures can often be used interchangeably (Wang, 2010). Efficiency measures assess the level of output for a given level of input (Martin & Frahm, 2010; Wang, 2010). Outcome or effectiveness measures assess the impact of the product or service on achieving the desired goals of the organization (Kapucu, 2010).

Although some scholars consider quality perspective within the outcome/effectiveness measures (Wang, 2010), the others specifically mentioned it as a separate dimension (Martin, 2002; Martin & Frahm, 2010; Martin & Kettner, 2010). In his report about the performance-based contracting practices of local governments, Martin (2002) argues that Governmental Accounting Standards Board (GASB) used expanded systems model (shown in Figure 1) to create a system of performance accountability, of which quality is an important perspective.

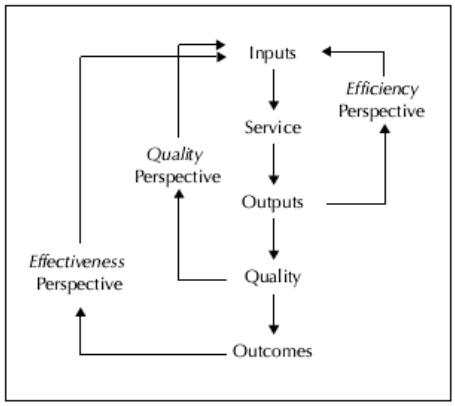

Figure 1. Expanded Systems Model

Wang (2002) notes that the use of a single measure is generally not sufficient for decision-makers; having more measures allows departments to illustrate a more complete overview of a program or service, thus better informing funding or policy actions. Moreover, it is also important to note that a small number of indicators may not suffice to reflect the complexities of some social phenomena (van Thiel & Leeuw, 2002; Perrin, 1998; Ingraham, Joyce, & Donahue, 2003). Holzer and Yang (2004) stress that the government agencies that are most productive are those who emphasize multiple measures, including internal capacities, outputs produced, and outcomes achieved.

In a similar vein, Ammons (2013) points out the importance of the refinement of performance measures in order to make them more meaningful and useful. As argued by Hubbard (2010), organizations often devote their time and energy to measuring things that

have little or no informational value and that are unlikely to contribute to managerial decisions, while neglecting to measure variables that have high informational value. Ammons (2013) stresses that governments should not only track their outputs but also develop measures focusing on efficiency, effectiveness, and service quality, which he calls as higher-order measures (Ammons 2002; Ammons & Riverbank, 2008). According to Ammons (2013), these higher-order measures "more often address desired results and either provide reassurance that performance is on track or sound an alert that it is not" (p. 511). Moreover, they are more likely than output measures to prompt managers, supervisors, and other staff to review whether the current performance is satisfactory, and, if not, what strategies should be devised to improve the performance (Ammons, 2002; Ammons & Riverbank, 2008).

Despite the widespread acceptance of the benefits of the higher-order measures, in practice, most governments utilize output measures much more frequently than these measures. Poister and Streib (1999) found that workload and output measures are the most frequently used measures in the US cities, whereas efficiency measures are those used less frequently. Further studies also supported these findings. de Lancer and Julnes (2001) found that 45% of responding state and local government representatives reported the use of output measures, while outcome (29%) and efficiency (24%) measures were reported less. This trend in the utilization frequency of the measures was again supported by Folz et al. (2009), in which 57, 50, 48, and 40% of the respondents reported the use of workload/output, quality, outcome/effectiveness, and efficiency measures, respectively, in their cities. This study also indicates that higher-order measures are more frequently used in the governments than reported in the earlier studies (Folz et al., 2009).

According to Bouckaert (1993), traditional performance measurement systems focus predominantly on the validity aspect, that is the technical problems of constructing a

good and a valid measure. However, optimal measurement systems should go beyond this one-dimensional system to include both legitimacy and functionality. The weakening of one dimension decreases the measurement capacity of the whole system and inhibits its potential benefits.

Another important issue regarding the performance measures is reviewing, revising, and updating the measures regularly. It is important keep the measures up-to-date. Out-of-date measures may not achieve their intended objectives. Moreover, sometimes problems or side effects some measures cause cannot be seen in advance and require a couple times of iterations of actual use. Reviewing and, when needed, revising performance measures after each performance measurement cycle contribute to the refinement of the measures (Perrin, 1998). Dixon, Nanni, and Vollmann (1990) emphasize the need for establishing a process that ensures performance measures to be reviewed as the organization's circumstances change. Similarly, Bititci, Turner, and Begemann (2000) point out the importance of performance measurement systems being dynamic in a way that reflects the changes in their external and internal environments.

Having explained above the literature about the performance measurement systems and performance measures, in the next part of the literature review, some of the predictors of effective performance measurement, which have been studied in the literature, will be examined in detail.

2.4 Predictors of Effective Performance Measurement

In the literature, several predictors of effective performance measurement have been mentioned and examined. In this part, stakeholder support for performance measurement and technical capacity of the organization regarding performance measurement will be discussed.

2.4.1 Stakeholder Support

In the literature, stakeholder support has been considered as one of the most important predictors of the effective performance measurement (Berman & Wang, 2000; Broad, 2006; de Lancer Julnes & Holzer, 2001; Fernandez & Rainey, 2006; Hatry, 2006; Taylor, 2006; Wang & Berman, 2001; Yang & Hsieh, 2007). Stakeholder support can be examined in two parts: external support and organizational support.

2.4.1.1 Organizational support. Organizational support is a frequently mentioned factor which has an important influence in the quality of performance measures and performance measurement systems. Regarding municipalities, organizational support concerns the support of mayor, managers, and employees. Organizational support for performance measurement can be considered closely related with the culture of the public organizations (Taylor, 2011). Whether the organizational actors have a managing for results culture can clearly affect their support for performance measurement initiatives (Melkers & Willoughby, 2005).

Support from mayor, top management, lower level management, and employees contribute substantially to the successful implementation of performance measurement systems (Berman & Wang, 2000; Melkers & Willoughby, 2005; Sanger, 2008; Poister & Streib, 1999; Streib & Poister, 1999; Yang & Hsieh, 2007). As put by Rosenberg (1998), level of readiness of the organization is fundamental for the success of any change initiative. Yang and Hsieh (2007) found in their study that organizational support is the most important predictor of effectiveness in both the adoption and implementation phases of performance measurement. In a similar vein, Folz et al. (2009) found that the lack of organizational support is the single most important factor that can explain why performance measurement fell short of meeting the expectations of chief executives.

There is a widespread agreement in the literature that any sort of management reform, including the introduction of performance measurement systems, will require support from top management (Berman & Wang, 2000; Denhardt & Denhardt, 1999; Fernandez & Rainey, 2006; Hatry, 2006; Hoontis & Kim, 2012; Kapucu, Volkov, & Wang, 2011; Wang & Berman, 2001; Yang & Hsieh, 2007). Top management support is important in the sense that it helps overcome resistance from lower level managers and employees, allocate budgetary and human resources, and maintain commitment for the performance measurement practices (Fernandez & Rainey, 2006).

Performance information has the potential to improve the managerial activities substantially (Behn, 2003). Yet the fear of being held accountable for results which they cannot entirely control can limit the support of mayors and top managers for these systems (Sanger, 2008). By using the Oliver's (1991) model of managerial responses to institutional pressures, Modell (2001) argues that the reactions of managers may range from supporting to compromising, avoiding, defying, or manipulating. Any strong reaction other than support can limit the effectiveness of performance measurement initiatives. In that sense, Fernandez and Rainey (2006) argue that it is important to first build high levels of commitment among top management and then get support from lower level managers and other staff.

However, the support of mayor and/or top managers alone is oftentimes not sufficient for the successful implementation of performance systems, which also requires support from lower level managers and employees. It is a well known and generally acknowledged fact that the resistance of lower level managers and employees against change may create significant challenges for top managers and sabotage the success of the performance systems (Berman & Wang, 2000; Poister & Streib, 1999; Streib & Poister, 1999; Taylor, 2006). Similarly, having acknowledged the benefits of employee support, Holzer and Yang (2004)

also mention the importance of employee participation in the design and improvement of performance measurement systems for enhancing the employee buy-in of these systems.

The positive role of organizational support in the success of performance measurement has been documented also by several empirical studies (Berman & Wang, 2000; Melkers & Willoughby, 2005; Sanger, 2008; Streib & Poister, 1999; Taylor, 2006; Yang & Hsieh, 2007). They mainly found that organizational support is an important predictor for the effectiveness of performance measurement systems.

In her study, in which she focused on the performance measurement systems in Hong Kong and Australia, Taylor (2006) examined to what extent these systems are valid, legitimate, credible, accessible, and functional. She found that measurement systems that receive the support of both higher and lower level employees are more likely to be better designed, implemented, and provide identifiable benefits for the organization (Taylor, 2006).

Berman and Wang (2000) also examined the role of organizational stakeholders on the use of performance measurement systems. County manager's office, department heads, managers, supervisors, and employees are among these stakeholders which have been included in the study. They found that the support of all of these groups is significantly associated with the level of use of performance measurement systems.

de Lancer Julnes and Holzer (2001) extended this study also to the implementation stage and examined the role of internal groups in both the adoption and implementation stages of performance measurement. However, their study also supported the findings of Berman and Wang (2000) in the sense that the support of internal groups is especially important in the adoption stage, not in the implementation stage. On the other hand, Yang and

Hsieh (2007) found that organizational support is the most important predictor of performance measurement not only in adoption level but also in implementation level (Yang & Hsieh, 2007).

2.4.1.2 External support. External support concerns to what extent are the performance measurement initiatives and practices being supported by the external stakeholders of the municipality. Elected officials and citizens are among the external stakeholders who are mentioned frequently in the literature (de Lancer Julnes & Holzer, 2001; Ho, 2006; Sanger, 2008; Wang & Berman, 2001; Yang & Hsieh, 2007). One of the most important contributions of the support of external stakeholders is that they allow or make it easier for the organization to allocate resources (de Lancer Julnes & Holzer, 2001) for developing and maintaining a technical capacity, which is necessary for developing high quality performance measures, and collecting and analyzing performance data. Another important benefit of external support is that it contributes to the utilization of information data even when the results contradict political agenda (de Lancer Julnes & Holzer, 2001).

As put by Zamesnik (2012), no discussion of performance measurement would be complete without mentioning the role that elected officials play in performance measurement. Support from elected officials is of critical importance for the implementation of performance measurement systems, since "it forecloses backchannels, legitimates reforms and new performance expectations, and helps ensure funding for new efforts" (Berman & Wang, 2000, p. 410). However, the literature points out the fact that elected officials do not always genuinely provide their support to the implementation performance measurement. It is claimed that elected officials are actually rarely interested in performance measurement, using it only as a tool to point out shortcomings with the current system, bureaucracy in general or a department in particular (Wang, 2002; Moynihan, 2008). Rather than a focus on

performance measurement as a tool to inform policy decisions and accountability, political figures view these measurement systems from the perspective of the impact they will have on their political agenda, constituencies, and reelection prospects (Hill & Lynn, 2009). Moreover, elected officials tend to regard performance initiatives from the executive branch with suspicion (Moynihan, 2008), and are not enthusiastic to support if they think that these initiatives are used as a tool by bureaucrats to avoid legislative scrutiny by technicalizing their operations (Kettl, 1994).

The general public's support for the use of performance measurement and their concern for the performance data are considered as important positive factors in the literature for the successful implementation of performance measurement systems (Berman & Wang, 2000; Ho, 2006; de Lancer Julnes & Holzer, 2001; Sanger, 2008; Yang & Wu, 2013). As argued by Berman and Wang (2000), support of citizens for performance measurement increases the legitimacy of the decisions and actions of both elected and appointed officials. Moreover, this support and interest in performance measurement and data create a pressure on these officials to use these data to communicate with the general public and to ensure a more efficient and effective delivery of public services (Ho, 2006).

Several empirical studies have examined the role of external support in performance measurement (such as Berman & Wang, 2000; de Lancer Julnes & Holzer, 2001; Yang & Hsieh, 2007). One of the early empirical studies about this issue is conducted by Berman and Wang (2000), who evaluated to what extent stakeholder support is associated with the use of performance measurement in the US county governments. They did not differentiate between external and organizational stakeholders, but they included stakeholders from these groups. As external stakeholders, they examined the role of elected officials, citizen advocates,

citizen advisory boards, and higher governments. They found that the support of these groups, except higher governments, is associated with the use of performance measurement.

In their study about US state and local governments, de Lancer Julnes and Holzer (2001) also examined the role of external stakeholders. They took into account the both adoption and implementation stages. As the external stakeholders, they included elected officials and citizens. They found that external stakeholder support is especially important for the implementation stage.

Yang and Hsieh (2007) also used the two-stage process of the utilization of performance measurement and examined the role of two external (political environment and stakeholder participation) factors in the adoption and managerial effectiveness of performance measurement systems. They collected data from government units in Taipei (capital city of Taiwan) and found that both factors are positively associated with the adoption and implementation stage of performance measurement (Yang & Hsieh, 2007).

2.4.2 Technical Capacity

In the literature, technical capacity of any organization is considered as one of the important factors of implementing successful performance measurement (Ammons & Rivenbark, 2008; Berman & Wang, 2000; de Lancer Julnes & Holzer, 2001; Streib & Poister, 1999; Wang & Berman, 2001). Some level of technical capacity is required for an organization to carry out its tasks and responsibilities (Wang, Hawkins, Lebredo, & Berman, 2012). According to Wang et al. (2012) building a capacity is a proactive action to motivate learning new behaviors to reinforce the implementation of new policies. Ingraham, Joyce, and Donahue (2003) points out the importance of capacity for the organizational performance.

Regarding the use of performance measurement, technical capacity refers to organizations' "ability to develop performance goals and measures and to overcome such conceptual barriers as distinguishing outcomes from outputs" (Berman & Wang, 2000, p. 410). In other words, technical capacity explains to what extent the organization can implement the performance measurement systems. This capacity includes both the human resources capability and the technological capacity of the organization. Although the technical capacity is essential for the successful implementation of performance systems, "the literature discussing specific technical competencies for performance measurement is surprisingly limited" (Berman & Wang, 2000, p. 410).

The quality of human capital is critical for the success of any organization. Without competent employees, even the best government policies cannot be implemented successfully (Kapucu, 2010). The same is valid also for the implementation of performance measurement systems. Developing performance goals and measures, collecting accurate and meaningful performance data, and analyzing these data require qualified human resources (Berman & Wang, 2000).

On the other hand, the deficiencies in the quality of human resources in this area may create a problem of measuring what can easily be measured, not what is meaningful to be measured (Moynihan, 2006; Hill & Lynn, 2009). Moreover, deficiencies in technical capacity may hinder the timely collection of performance data, which is important for the success of performance measurement (Fryer at al., 2009). These problems, in turn, may limit the possibility of achieving the intended benefits of performance measurement systems. To overcome this problem, the strategy and practice of recruiting, and training human resources are very central to the effectiveness of the performance measurement systems. However, recruiting qualified personnel, and training them requires a significant financial investment

for organizations, and this may create problems especially for those organizations with limited resources (Zamesnik, 2012). The support of stakeholders, especially those of elected officials and managers are of special importance to overcome or mitigate such problems.

Another factor considered within the technical capacity is the technological infrastructure of the organization. Specifically, the collection and analysis of performance data require an important investment in information technology, which has always the potential of creating problems for those organizations, which have limited resources especially in the times of substantial budget constraints (Zamesnik, 2012).

Technical capacity is also important for conducting citizen satisfaction surveys which have become quite popular as a source of performance measurement (Taylor, 2006). Although the information gathered through the surveys are related to perceptions and therefore subjective, they have the ability to directly give information about the citizen satisfaction with the public services, which can be used to improve the services (Holzer & Yang, 2004). However, as argued by Berman and Wang (2000) conducting citizen surveys in scientifically valid ways can be costly, both in terms of technology needed to collect and analyze the data, and in the cost of administering the actual survey, and many jurisdictions do not have this capacity. Contracting for such information may be an alternative, but it also requires substantial monetary resources and may create problems with the timely gathering of information (Berman & Wang, 2000). Additionally, there may be issues with compatibility of a new data reporting system with existing information technology (IT) programs, and issues with a timeline for implementation of a new system, including training of employees and troubleshooting unexpected issues. Having qualified human resources and necessary technical infrastructure may ease these problems.

Berman and Wang (2000) examined the role of technical capacity in the adoption of performance measurement in the county governments and found that technical capacity is strongly associated with the increased use of performance measurement. In a follow up study, they also examined the association between professional competence and deployment of performance measurement, including the use of output and outcome measures (Wang & Berman, 2001). They found a positive association between the professional competency and the use of both output and outcome measures (Wang & Berman, 2001).

2.5 Theoretical Framework

In the study, context-design-performance (CDP) model, or the context-design theory, as called by Wang (2010), is used as the main theoretical guidance to examine the predictors of effective performance measurement in Turkish municipalities. This model is heavily influenced from contingency theory (Agiro, 2011; Burke & Litwin, 1992; Goltz, 2006; Marathe, 2006; Wan 1995) and Donabedian's structure, process, and outcome model (Agiro, 2011; Goltz, 2006; Wan, 2002). In the following sections these theoretical approaches and how they are useful in guiding this study will be explained.

2.5.1 Context-Design-Performance Model

Context-design-performance model is one of the theoretical frameworks that are useful to specify the causes of organizational performance. The basic model of CDP is shown in Figure 2. In this model, contextual factors influence organizational performance both directly and indirectly (via design factors). Other than the contextual factors, design factors also influence organizational performance (Wan, 1995). As opposed to the production process theory (input – process – output – outcome) or Donabedian's Structure, Process, and Outcome (SPO) model, this model does not focus only on an organization's internal

operations and it is especially applicable when external factors play an important role in the performance of organization (Wang, 2010).

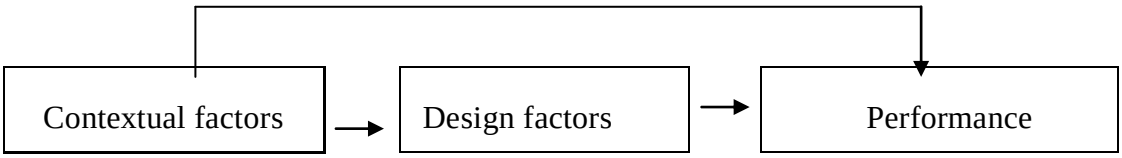

Figure 2. Context-Design-Performance Model. Adapted from Wan, 1995.

This model is an adaptation of a specific model (context – structure – performance) (CSP) of structural contingency theory (Agiro, 2011) and Donabedian's Structure, Process, and Outcome model to the organizational level (Wan, 2002). As a result, it combines both the external and internal factors which influence the organizational performance.

In this model, context factors include characteristics of environment, organizational culture, technology, or size of organizations (Drazin & Van de Ven, 1985). As a result, contextual factors can be external or internal factors (Lin & Wan, 1999; Mark, Salyer, & Wan, 2003). Design factors in this model exceed the limits of organizational structure in CSP model and include measures of capability and capacity, which overlaps more with the notion of structure in Donabedian's SPO model (Agiro, 2011). Performance can be conceptualized by several ways such as including outcomes, efficiency, productivity, quality, and effectiveness (Flood, Zinn, & Scott, 2006).

Wan (1995) used the CDP model in explaining the components of health care delivery systems. This model was mostly used in the health care field (Agiro, 2011; Lin & Wan, 1999; Mark et al., 2003). Yet as an exception, Goltz (2006) used this model to examine the role of environmental and design factors in the organizational performance of police organizations in Florida.

After explaining the CDP model, two theoretical approaches which lay the foundations for this model will be explained in the following parts.

2.5.2 Contingency Theory

Early universalistic theories of organizations focused on the internal elements of an organization and sought for the one best way to organize (Donaldson, 2001). In response to these theories, contingency theories, which emphasized the importance of environmental factors in shaping an organization's structures, emerged during the 1950s (Weill & Olson, 1989) and gained popularity in the 1960s (Maguire, 2003). Contingency theory has dominated the study of organizational design since the mid-1960's (Scott, 2003) and is still "the most widely utilized contemporary theoretical approach to the study of organizations" (Scott & Davis, 2007, p. 104). Contingency theory is an important open system theory, which emphasizes the importance of external factors on organizational performance, as opposed to closed system theories which focus on internal operations. Different from classical management theories, which try to find the best way for organizations to be operated and managed, contingency approach proposes that there is no best way which is valid for every organization (Donaldson, 2001; Scott & Davis, 2007).

Donaldson (2001) defines contingency as "any variable that moderates the effect of an organizational characteristic on organizational performance" (p. 7). There are contingency theories of many different organizational characteristics, such as organizational structure, design, leadership, strategic decision making processes, and human resources management (Donaldson, 2001; Scott, 2003). In general, "contingency theory is guided by the general hypothesis that organizations whose internal features best fit the demands of their environments will achieve the best adaptation" (Scott, 2003, p. 96).

Scott (2003) points out the importance of design decisions in contingency theory and emphasizes that according to this theory, design decisions are contingent upon environmental factors. In his book, where he focused only on the structural contingency theory, Donaldson (2001) mentions three core elements which form the core paradigm of structural contingency theory, which can be more or less considered as valid for all contingency theories if the structure is replaced by the internal feature on which that contingency theory focuses. The first one is the presence of an association between contingency and organizational structure. Secondly, since a change in contingency requires a change in structure, it can be concluded that contingency determines the organizational structure. Thirdly, the fit between contingency factors and organizational structure "leads to higher performance, whereas misfit leads to lower performance" (Donaldson, 2001, p. 7). According to Donaldson (2001), and Drazin and Van de Ven (1985), this fit-performance relationship lies at the center of contingency theory approaches. However, there has not been a consensus about the definition of fit or match (Mark et al., 2003).

There are several studies which used the contingency theory approach in the field of performance measurement (Bititci, Turner, & Begemann, 2000; Dixon, Nanni, & Vollmann, 1990; Klovienė, & Gimžauskienė, 2008; Rejc, 2004). Dixon et al. (1990) emphasize the need for establishing a process that ensures performance measures to be reviewed as the organizations's circumstances change. Similarly, Bititci et al. (2000) point out the importance of performance measurement systems being dynamic in a way that reflects the changes in their external and internal environments.

Lenses of contingency theory are used in the study in order to examine the external factors which have influence on the organizational performance. However, contingency theory does not suffice alone as a theoretical framework in this study, since it assumes all

contingencies are equally important in influencing the performance of organizations (Agiro, 2011).

2.5.3 Structure, Process, and Outcome Model

Donabedian's SPO model is a helpful and prominent model in examining the organizational factors which influence organizational performance. It is mostly used in health services administration research (Flood et al., 2006). The main idea of the model is that good structure is expected to promote good process and good process in turn is expected to promote good outcome (Zinn & Mor, 1998; Donabedian, 1988). As correctly put by Goltz (2006), Donabedian's linear relation can be considered as "a simplified version of a much more complex reality" (p. 17).

Although contingency theory is useful in emphasizing the importance of external factors on the organizational performance, it fails to examine the relative importance of the determinants of organizational performance (Agiro, 2011). Moreover, contingency theory focuses on the fit between environment and structure or other internal features. However, fit of contingencies may not be enough to examine to what extent the external and internal factors are important in the organizational performance (Agiro, 2011). In that sense, SPO model, which is mainly a systems theory model, in which "input-throughput-output with a feedback loop is the basic model" (Burke & Litwin, 1992, p. 524), helps for a more precise analysis of the factors which have direct and consistent influence on performance (Agiro, 2011).

2.6 Conceptual Framework

Since expected benefits of performance measurement can only be ensured with the successful implementation, it is important to examine the factors which affect the effective implementation of performance measurement systems in public agencies. In the literature, external and organizational support for the use of performance measurement and technical capacity of organizations for performance measurement have been noted as some of the most important factors which influence the implementation of performance measurement (Berman & Wang, 2000; Broad, 2006; de Lancer Julnes & Holzer, 2001; Fernandez & Rainey, 2006; Hatry, 2006; Poister & Streib, 1999; Streib & Poister, 1999; Taylor, 2006; Wang & Berman, 2001; Yang & Hsieh, 2007). This study included these factors as the predictors of effective performance measurement. In this study, external support and organizational support have been considered as the contextual factors, and technical capacity as the design factor. The proposed conceptual model of the study is given in Figure 3.

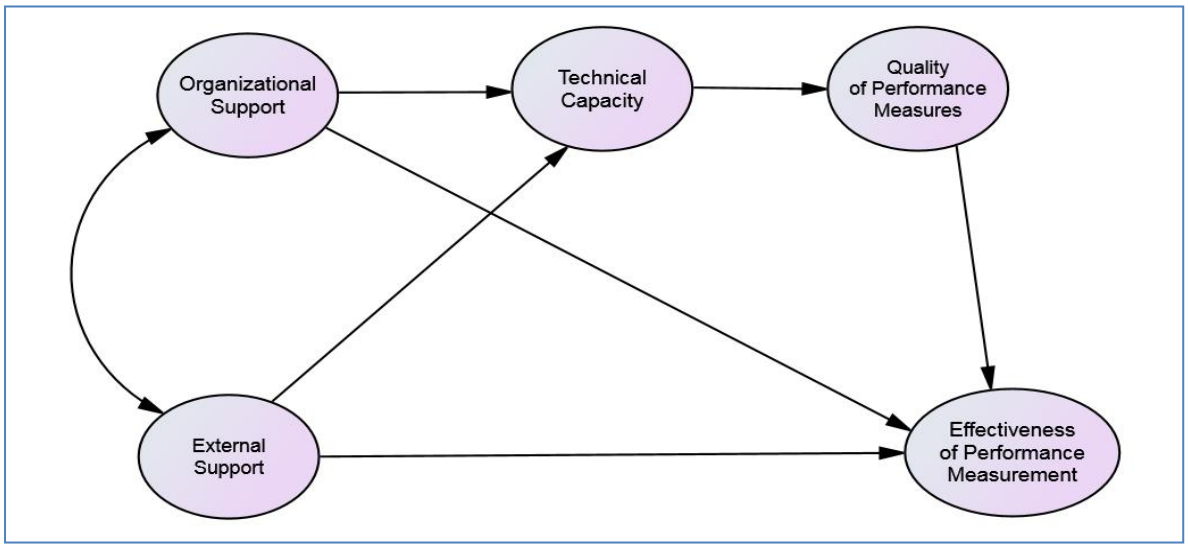

Figure 3. Conceptual Model of the Study

According to de Lancer Julnes and Holzer (2001), utilization of performance measurement systems consists of two stages: adoption of performance measures and

implementation of performance measurement. As argued by Yang and Hsieh (2007), it is possible that adoption and implementation stages of performance measurement may be affected by different factors. In accordance with that, quality of performance measures, which is the performance of organization in adoption stage, and effectiveness of performance measurement, which is the performance of organization in implementation stage, are considered as variables of performance in this study.

The relationships among the study variables, which are shown in Figure 3, are the hypotheses of the study and will be explained in the following parts in detail.

2.6.1 Effectiveness of Performance Measurement Systems

In general, effectiveness of performance measurement concerns the extent to which performance measurement achieves its intended results and objectives. Performance data can be used in areas such as strategic planning, decision-making, budget allocations, and communication to internal/external stakeholders. Moreover, performance measurement can contribute to the managerial goals by improving productivity (Behn, 2003; Hatry, 2006) and service quality, increasing employee motivation, and stimulating organizational learning (Yang & Hsieh, 2007).

The use of performance measurement system does not necessarily guarantee the achievement of its intended objectives. In their survey of approximately 300 local government administrators, Melkers and Willoughby (2005) found that almost 70% of local governments in their sample use performance measurement in at least half of their departments. However, the respondents reported important deficiencies in the "effectiveness of using performance measures to influence budgeting processes and outcomes in particular" (Melkers & Willoughby, 2005, p. 188). Similarly, de lancer Julnes and Holzer (2001) found

that only a smaller set of local governments that measure their organizational performance actually use them to improve their managerial decisions.

2.6.2 Quality of Performance Measures

Development and adoption of high quality performance measures is one of the important prerequisites of implementing a successful performance measurement system. This study uses Bouckaert's (1993) model in assessing the quality of performance measures. According to this model, performance measures can be considered as effective, if they are valid, legitimate, and functional. In this model, validity refers to the technical soundness of performance measures. In other words, in order to be considered as valid, performance measures should be "sound, cogent, convincing, and telling" (Bouckaert, 1993, p. 31). Streib and Poister (1999) also stress the importance of understandability and "the need to base measures on a mission statement and clear objectives" (p. 109) in evaluating the validity of performance measures. In a similar vein, Sanger (2008) argues that "measures must be clear, accurate, and credible to internal and external audiences" (p. S80).

Legitimacy of performance measures is about the perceptions of stakeholders regarding the performance measures. According to Bouckaert (1993), performance measurement is not only a technical issue, but also a motivational one. Involvement of employees and middle managers in the creation of performance measures can increase their commitment to performance measurement, effective implementation of which requires the approval of these groups (Bouckaert, 1993).

Finally, functionality of performance measures refers to the benefit creating potential of the measures. Regarding this dimension, Bouckaert (1993) stresses on that performance measures should contribute to the maintenance or to the development of the organization.

Otherwise, they can even become dysfunctional by causing behaviors that contradict the intended goals or the purposes of the organization.

This model was used by Streib and Poister (1999) in their study, in which they evaluated the design of performance measurement systems of the US municipalities with populations more than 25,000. They collected data by a survey sent to senior officials of municipalities. They found that despite some municipalities which perform fairly well in the sense of validity, many others still tend to focus only on available data, and measuring the quality of the services they provided seems to be an especially hard task for them. Regarding legitimacy, they found that involvement of lower level employees and citizens in the design of performance measures is a very rare event. Pertaining to functionality dimension, they stated that, other than manager accountability and employee focus on organizational goals, the benefits accrued appear to remain limited in their sample.

In their important study, de Lancer Julnes and Holzer (2001) found that the level of adoption has an important influence on the level of implementation. This association is also confirmed later by Yang and Hsieh (2007). In this study, the performance of the organization in the adoption stage of performance measurement is conceptualized as the quality of performance measures, whereas performance in the implementation stage is conceptualized as the effectiveness of performance measurement. Based on the arguments of these authors, it is hypothesized in this study that:

H_1: Quality of performance measures is positively related with the effectiveness of performance measurement systems.

2.6.3 Technical Capacity

In the literature, technical capacity of the organization is considered as one of the important factors of implementing successful performance measurement (Ammons & Rivenbark, 2008; Berman & Wang, 2000; Streib & Poister, 1999). Like many other government policies, performance measurement cannot be effectively implemented without qualified human capital and necessary technical infrastructure. Developing performance goals and measures, collecting accurate and meaningful performance data, and analyzing these data require qualified human resources and sufficient technical infrastructure (Berman & Wang, 2000). Based on these explanations, it is hypothesized in the study that:

H_2: Technical capacity of the municipality influences the effectiveness of performance measurement indirectly via the quality of performance measures.

2.6.4 External Support

External support has the potential to contribute to an increase in the technical capacity for performance measurement, since resources related to capacity building in this area is highly dependent on the approval of elected officials. Moreover, citizen support both creates a pressure on the development of capacity and also increases the legitimacy of decisions made in this regard (Yang & Wu, 2013). However, even though performance measurement, with the help of increased technical capacity and adoption of high quality performance measures, creates accurate and useful performance data, this may not still be a sufficient factor for the utilization of these data. Several studies showed that most of the important decisions, especially the ones with budgetary results, are made with political incentives rather than objective/rational criteria (de Lancer Julnes & Holzer, 2001). Therefore, it can be argued that external support for performance measurement has also a direct effect on the

effectiveness of performance measurement. Based on these explanations, it is hypothesized in the study that:

H_3: External support for performance measurement influences the effectiveness of performance measurement indirectly via technical capacity of the municipality and the quality of performance measures.

H_4: External support for performance measurement directly influences the effectiveness of performance measurement.

2.6.5 Organizational Support

There is a widespread agreement in the literature that organizational support for performance measurement contributes substantially to the successful implementation of performance measurement systems (Berman & Wang, 2000; Melkers & Willoughby, 2005; Sanger, 2008; Streib & Poister, 1999; Yang & Hsieh, 2007). Even some authors found that organizational support is the most important predictor of effectiveness in both the adoption and implementation phases of performance measurement (Yang & Hsieh, 2007).

In this study, organizational support is examined as the support of mayors, top managers, lower level managers, and other employees. Although mayors are elected officials, their support is considered as organizational support not external support, since Turkish municipalities are governed by a strong mayor government system. The mayor is the head of municipal administration and he/she can directly intervene in every decision made and activity done by the departments of the municipality.

Organizational support for performance measurement can contribute to the development of high quality performance measures by providing required resources for the development of technical capacity and also by motivating the personnel about the importance

of performance measurement. However, if performance measurement initiatives are adopted only because of external pressures and/or institutional isomorphism, and if they are not supported by organizational actors, this may create challenges for the effective implementation of performance measurement (Moynihan, 2005).

Moreover, in the literature, it was argued that there is a positive relationship between the levels of organizational support and external support (Yang & Hsieh, 2007). It is possible that the support from citizens and council members may affect how organizational actors perceive performance measurement. On the other hand, organizational actors that support the use of performance measurement can impact the perceptions of external actors by emphasizing the benefits of performance measurement in every occasion. Since the relationship may go in both ways, a covariance among these variables is hypothesized in the study. Based on these explanations, it is hypothesized in the study that:

H_5: Organizational support influences the effectiveness of performance measurement indirectly via technical capacity of the municipality and the quality of performance measures.

H_6: Organizational support directly influences the effectiveness of performance measurement.

H_7: There is a positive correlation between organizational support and external support.

CHAPTER 3: THE CONTEXT OF THE STUDY

In this part, background information about the Turkish public administration, administrative reforms, and local government systems will be given in order to allow readers a better understanding of the context of the study.

3.1 Turkish Public Administration and Administrative Reforms

Every country in the world experiences administrative reforms regardless of their government styles, political/administrative culture or development levels (Sezen, 2011). Turkey is not an exception to this case and administrative reforms have always been an important agenda of political life. The Turkish Republic, which was founded in 1923, inherited a highly centralized and bureaucratic state from the Ottoman Empire (Ozcan & Turunc, 2008). Consequently, the Turkish state has historically been a more dominant actor than the civil society in Turkey (Sozen & Shaw, 2002).

Turkey experienced the most comprehensive administrative reforms in the early Republican period. In this period, the Ottoman political and administrative heritage was completely transformed (Berkman & Heper, 2002; Sezen, 2011). The Republic of Turkey, which has been based on the principles of the parliamentarian democracy, secular social regime, unitary state, and administrative jurisdiction, established its main legislation and institutions during the 1920s and 1930s on the model of the Western world. Regarding administrative structure, mainly the French administrative system has been adopted, which has resulted in a highly centralized structure (Celenk, 2009; Sezen, 2011). This centralized aspect is also apparent in the Municipality Law of 1930, which saw the municipalities as an extension and representative of the central government. The law established an administrative tutelage, according to which some of the important decisions of municipalities required the

approval of local branches of central administration. The law also created a uniform administrative structure for the municipalities, which will be applied to the whole country, regardless of geographical, cultural and economic differences. Moreover, in these years, the mayors were appointed and almost all of the financial resources of the municipalities were provided by the central government (Celenk, 2009).

Besides the early period in the foundation of the Republic, two main sets of administrative reforms were carried out in the Turkish administrative system. Both of these radical sets of reforms, which included also the adoption of new constitutions, took place after the military coups in 1960 and 1980. These reforms significantly changed both the Turkish political and administrative systems, including the tasks and responsibilities of local governments (Sezen, 2011).

In 1960, the last Democrat Party government was overthrown by the military officers. Before holding new elections in 1961, the military regime introduced a host of reforms, including a new constitution. These reforms aimed to lay the foundations of the welfare state, an important part of which was the transition to the planned economic and social development, in Turkey (Berkman & Heper, 2002). The mayors have become elected by the citizens for the first time in 1963, however, its role in decreasing the central control over the local governments remained limited (Celenk, 2009). The transition to the planned economy did even increase the centralist pressures regarding local public services, since this regime was inclined to centralize decisions, resources, and tasks with aim of enabling the implementation of macro socio-economic plans (Bayraktar, 2007).

In September 1980, the military took power into its own hands for the second time. This time, too, intervention was followed by a series of reform programs, including administrative ones (Berkman & Heper, 2002). The reforms following the 1980 coup and

their effects on Turkish administrative system still prevail, since the constitution which was adopted in 1982 is still in effect. This main crux of the reforms in this period is the dissolution of the welfare state, which changed the development strategy and the role of the state in planning the economical and social development (Sezen, 2011). The Motherland Party, which captured power following the 1983 general elections, adopted a policy that replaced the earlier economic policy of import substitution in particular and *e´tatism* in general, with the export-promotion and privatization (Ozcan & Turunc, 2008), which meant reducing the scope of civil bureaucracy in Turkish politics and economics (Berkman & Heper, 2002).

According to Bayraktar (2007), the main aim of this neo-liberal trend was mitigating the burden of the state, both administratively and financially. In accordance with this perspective, local authorities were seen as important bodies, which can alleviate the central tasks and responsibilities. Consequently, financial resources and administrative capacities of Turkish municipalities have begun to enjoy a gradual and steady improvement for the first time (Bayraktar, 2007). Other than the financial improvements, local governments have also experienced an increase in their powers, tasks, and responsibilities (Bayraktar, 2007; Ozcan & Turunc, 2008).

Although the 1980s and 1990s experienced a momentum of economic reforms, administrative reforms were only partially carried out. Finally, the Justice and Development Party (AKP), which came into power in 2002 and is still in power, has radically transformed the Turkish public administration. The AKP came to power in 2002 with a comfortable majority and have maintained their power in the 2007 and 2011 elections with even bigger majorities. According to Sezen (2011), having this strong government is one of the most

important reasons which made it possible for the AKP to conduct these reform policies in this period.

Because of the administrative tradition and culture developed from and shaped by the previous Ottoman Empire regime, which was highly centralized and bureaucrat- dominated, the current public administration system in Turkey includes substantial red tapes, organizational inefficiency and ineffectiveness, misuse in the use of resources, and consequently the inability to meet public needs (Kapucu & Palabiyik, 2008; Kosecik et al., 2003). To overcome the problems faced in the Turkish administration system, several research projects and reform initiatives were conducted. However, none of them created significant results, which can contribute substantially to overcome the social, economic, and political problems (Kapucu & Palabiyik, 2008). According to Kapucu and Palabiyik (2008), the reforms conducted prior to the end of the 20th century were basically a repetition of each other and provided only superficial solutions to the problems. However, at the beginning of the 21st century, in response to both its domestic demand on reforms in economic, social, administrative, and political reforms and in particular the requirements stipulated by European Union (EU) membership standards, Turkey has launched important efforts of restructuring its reforms to build up governance capacity (Kapucu, 2010).

In this period, the EU has played a significant role in the introduction of reforms (Bayraktar, 2007; Celenk, 2009; Guney & Celenk, 2010; Ozcan & Turunc, 2008; Sezen, 2011; Sozen & Shaw, 2002). Approval of the candidacy status of Turkey for the EU membership in 1999 and the beginning of the accession negotiations in 2005 increased significantly its importance in the transformation of Turkish public administration. The reforms done in this period to comply with the EU regulations have created significant social, economic, and political changes in Turkey (Sozen & Shaw, 2002).

Other important institutions, which have had an important influence upon Turkey's political and economic policies, are some key financial organizations, such as the World Bank (WB) and the International Monetary Fund (IMF) (Sozen & Shaw, 2002). Turkey was dependent on international loans since the 1940s and as a result, was subject to insistent demands emanating from international agencies since the beginning of the 1970s. Because of these pressures, the country changed its development strategy substantially in the 1980s and a series of public administration reforms have been undergoing since then. Through the loan agreements and letters of intent, the WB and the IMF do not only demand modifications on economic and financial policies, but also administrative reforms which are necessary for the implementation of these policies (Sezen, 2011). As a result, it can be concluded that the EU and the other international organizations motivated the Turkish governments to implement administrative reforms, which are influenced by the reforms movements explained in the previous sections.

According to Kapucu and Palabiyik (2008), the new tide of administrative reforms, which started in 2003 and is still ongoing, claims to be different from all the other previous reform attempts in Turkey. These reforms are built on fundamental values that depend on good governance, such as management in place, respect to human rights, accountability, transparency and the effective use of resources. Along with this, in the organization and operation of the public administration, strategic management, and performance management are taken into consideration. The Law on Basic Principles and Restructuring of Public Administration, which is the legal text of the reform, had been accepted by the Turkish Grand National Assembly (TGNA) in 2004, but it was vetoed by the President and so it was not put into practice. Nevertheless, the complementary legal regulations relating to the basic goals and targets of the reform have been tried to be achieved step by step. In that sense regulations

such as; the Public Finance Management and Control Law (2003), The Freedom of Information Law (2003), The Municipality Law (2005), The Greater City (Metropolitan) Municipality Law (2004), and The Special Provincial Administration Law (2005) was adopted to bring about the change in the direction of the new perspective in the public administration (Kapucu & Palabiyik, 2008).

Research on public administration, academic literature, and government programs in Turkey generally assert that the central government is abusively strong, whereas the local governments are very weak and lacking financial and administrative autonomy. Therefore, there is a necessity to strengthen the local governments for the sake of democracy (Bayraktar, 2007; Sezen, 2011). In that sense, the new laws about local governments are especially important, since they reversed the distribution of tasks and responsibilities between central and local governments in favor of the latter. In the previous system, the tasks and responsibilities of local governments were listed and all other tasks were considered as the responsibility of the central government. With the new system this distribution has been reversed by the favor of the local governments by restricting only the tasks of central government and leaving all other to the local governments (Sezen, 2011). With this new distribution, many tasks and responsibilities which were carried out previously by the central government have been transferred to the local governments. These reforms also weakened the administrative and financial control of the central government over the local governments, while enhancing their autonomy (Bayraktar, 2007; Sezen, 2011).

Another important reform for local governments and performance measurement is the Public finance management reform, which aims to control and reduce the public expenditure. The Public Finance Management and Control Law, adopted in 2003, completely reorganized the public finance management system, which is in operation since 1926. The new system

requires public agencies to prepare their budgets for a three-year period in accordance to their strategic plans including performance measures. Moreover, managers of public agencies must clearly show the extent that they achieved the targeted performance in their annual reports (Sezen, 2011).

Strategic plans have also become compulsory for local governments with more than 50,000 dwellers. Strategic plan is prepared by the mayor, and acquires legal status after it is approved by the council. Strategic plan and performance evaluation program are very critical in budget preparation; that is why they are discussed and approved in the council before the budget (Kapucu & Palabiyik, 2008; Karasu & Demir, 2012). However, Sezen (2011) argues that, with the exception of metropolitan municipalities, most of the municipalities do not have a planning tradition and qualified personnel for planning. Most of the municipalities prepare strategic plans only to fulfill the regulatory obligations, not to guide their upcoming works effectively. Sezen (2011) explains these problems by pointing out the difficulties in transforming the political and administrative culture in the short term only by passing laws. Issuing a law could just be a beginning; the harder issue is to enforce the law. Turkey has sufficient experience of many unimplemented reformist laws (Kapucu & Palabiyik, 2008). As correctly put by Sezen (2011), "[e]ven if it is easy to adopt [such laws] formally, they are either not put into effect or they are just a façade" (p. 340).

3.2 Turkish Public Administration

Turkey is a centralized and a unitary state governed by a parliamentary democratic system. It has a highly centralized administrative structure. The administration of Turkey is composed of central and local administrative agencies (Kapucu & Palabiyik, 2008). In the following parts, service areas and structure of the municipalities will be explained in detail within the context of Turkish public administration. Turkish administrative structure is shown

in Figure 4. Regarding central administration, local branches of government will receive additional emphasis, because of their close working relationship with the municipalities.

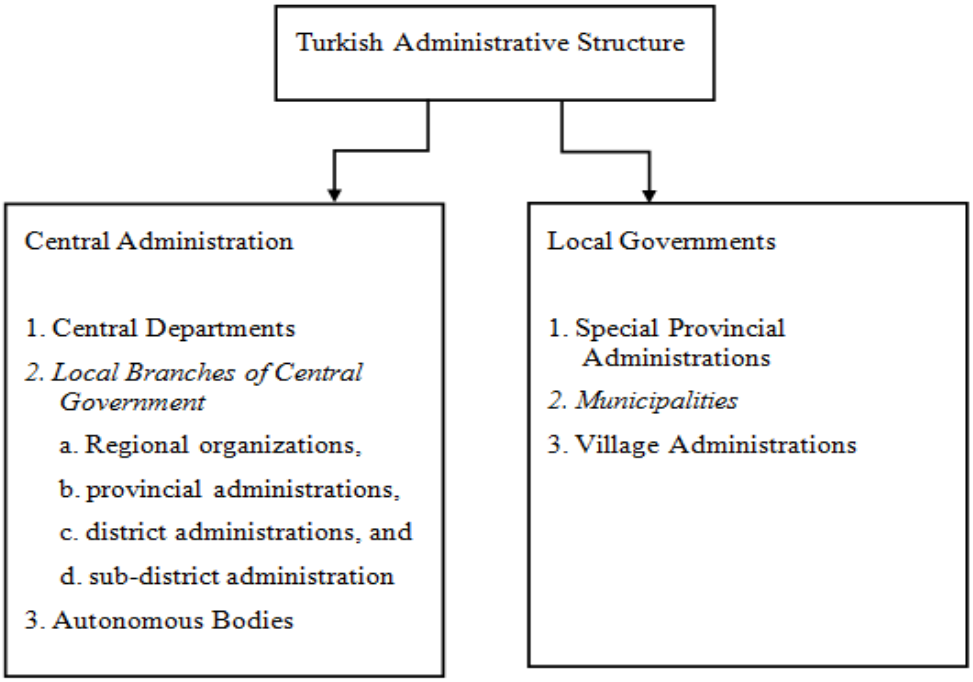

Figure 4. Turkish Administrative Structure. Adapted from United Nations Public Administration Network (n.d.)

3.2.1 Central Administration

The structure of central administration comprises of central state apparatus: prime ministry, ministries and other related government agencies and their local branches in provinces and districts. Autonomous bodies are not of interest for this study. The local branches of the central administration can be categorized as the following: regional organizations, provincial administration, district administration, and sub-district administration. These branches provide services in the name, and in line with the imperatives and instructions, of the central administration. Since regional organizations are exceptions and are only limited for some specific services, and the sub-districts are, in practice, facing

extinction with no more appointment to their posts, this paper focuses only on province and district level regarding the local branches of central administration.

Turkey is divided into geographic regions called provinces. There are 81 provinces in Turkey, which are divided into 919 districts (Turkish Ministry of Interior, 2014a). The main local administration branch of the central administration is province. The provinces are governed by governors representing and responsible to the central government, by which they are appointed. The governor represents the legal personality of the state as well as of each ministry in provinces separately. As an agent of central administration, the governor has substantial authority on local branches of central administration and their staff, in addition to being head of law enforcement agencies in provinces. The governor maintains harmony between central and local government services. Every ministry has its headquarters in the provinces, and above all of the respective ministries are the provincial administrators. Provincial administrators are appointed by the respective ministry and receive orders from, and are responsible to, the governor.

The district is a subordinate agency of the central administration and governed by the district governors, which are also appointed by the central government and represent the government. The district governor executes the orders and directives of the provincial governor. The district branches of the ministries are administered by the district administrators who receive orders from and are responsible to the district governor.

3.2.2 Local Governments and Municipalities

Unlike local branches of central administration, local governments are democratic administrative units with certain degree of autonomy in terms of financial and administrative issues; they function outside the central administration to provide common and local services,

and are governed by decision-making organs, which are directly elected by the people (Kapucu & Palabiyik, 2008). There are three types of local governments in Turkey: municipalities (including metropolitan municipalities), special provincial administrations (SPA), and villages. SPAs are local governments which are established (one per province basis) to carry out tasks in the areas that fall neither within municipal or village boundaries. Villages are small settlements consisting of usually fewer than 2,000 inhabitants. Since this study concerns only municipalities, SPAs and villages are excluded in the following parts.

Municipality is a form of local government established to meet the local needs of the urban residents. According to the Municipal Law, municipalities are established at the centers of provinces and districts whatever their populations are, and in the settlements of whose population is more than 5,000. Municipalities are administratively and financially autonomous corporate public entities with legal personality. Decision making organs are formed through local elections to provide common local services assigned to them within their jurisdictions. According to the Municipal Law (Article 14) municipalities can provide local and common services such as urban planning; water and sewage systems; transportation; environment and environmental health, sanitation and solid waste; fire department, emergency aid; city traffic; forestry, parks and recreation; housing, culture and art, tourism; social services; women and children shelters, supporting education, health, and sports in their jurisdictions. Moreover, as explained above, the municipalities may carry out other duties which are not in the responsibility areas of other public agencies.

Municipality has three main organs: municipal council, municipal executive committee, and mayor. The mayor is the head of municipal administration and is elected directly by the citizens through local elections. Among the duties and the powers of the mayor are: to direct and manage the municipality in accordance with the strategic plan; to

prepare the budget, and to determine, monitor, and assess the performance measurement of municipal activities and personnel on the basis of these strategies; to submit an activity report to the council; to chair the council and executive committee meetings, and to execute decisions taken by them.

The council is the main decision-making organ of the municipality. The council takes most of the important decisions, such as the approval of the strategic plan, investment and work programs, performance scale of the personnel, budget and final accounts, and development plans. The members of the council are directly elected by the local citizens. The municipal council is chaired by the mayor. The monthly agenda of the council is prepared by the mayor. On the other hand, the council assesses the work of the municipality through the annual activity reports. If the activities in the report are found to be unsatisfactory by ¾ of the members of the municipal council, the governor is informed of the dissatisfaction. The governor sends the case to the Council of State with his/her reasoned opinion. If Council of State decides incompetency of the mayor then the mayor is unseated. Moreover, one third of the council members may propose an interpellation of the mayor, which is finalized by the above-explained method.

The municipal executive committee is considered as both the decision-making and the counseling organ of the municipality. It presents its comments to the council regarding the strategic plan, budget, final accounts, and annual work program. It also takes some important decisions regarding the functioning of the municipality, which are articulated in the Municipal law and other laws. The municipal executive committee is led by the mayor, and is composed of some council members elected by the council and by some municipal administrators appointed by the mayor.

Metropolitan municipalities are considered as a special form of government for cities which have more than 750,000 people residing within its borders or within 10,000 meters (approximately 6.2 miles) around its borders. Metropolitan municipality can be defined as "a municipality which has more than three district or lower-tier municipalities within its boundaries" (Kapucu & Palabiyik, 2008, p. 115). As can be understood by the definition, a two-tier system is considered in metropolitan cities (Kapucu & Palabiyik, 2008). There were 16 metropolitan municipalities in Turkey until recently. However, the laws regarding the establishment of 14 more metropolitan municipalities passed from the TGNA and they have become effective with the recent local elections, which were held in March 2014. Moreover, with these laws the jurisdictions of the metropolitan municipalities are widened to encompass all the provincial territories including the rural areas.

Compared to the other municipalities, metropolitan municipalities have more budgetary and human resources (Kapucu & Palabiyik, 2008). On the other hand, the organs of metropolitan municipalities and their interactions with each other are more or less similar to those of other municipalities. The metropolitan municipal council is the main decision-making organ of the municipality. It is formed with the participation of one fifth of all district or lower-tier municipal council members in local elections. Mayors of district and lower-tier municipalities are natural members of the metropolitan municipal council. The metropolitan executive committee is both a decision making and executive organ, and also an advisory committee of the metropolitan municipality. It consists of five metropolitan council members, elected by the council, and five department supervisors, appointed by the metropolitan mayor. Both the council and the executive committee meetings are chaired by the metropolitan mayor. The mayor, who is elected directly by the local citizens, is the head of metropolitan administration and representative of its legal personality.

CHAPTER 4: METHODOLOGY

This section of the study provides the methodological information that shows how the research was carried out. In this study, quantitative research methods (Creswell, 2009), such as structural equation modeling, were utilized to analyze the data which was collected via a self-administered online survey. The study used a non-experimental single group research design. This study is an organizational level study and as a result, all the data collected for the study are at the organizational level. Unit of analysis for the study is municipality. In this chapter, study variables and their operationalization, data collection, statistical analyses, measurement models, and their validation will be explained in detail.

4.1 Study Variables

The study has five latent variables, two of which are exogenous variables (organizational support and external support). Technical capacity, quality of performance measures, and effectiveness of performance measurement are endogenous variables, first two of which are also mediating variables. Contextual explanations of these variables and related literature have already been explained in the literature review and conceptual framework sections of this study. In this part, the operationalization of these variables will be illustrated.

The study has also two control variables: population and the type of the municipality. In the literature, population has been considered to be related with the utilization of performance measurement (Folz et al., 2009; Riverbank & Kelly, 2003). It is also possible that effectiveness of performance measurement may vary across the types of municipality, since they have different administrative structure and resources. Detailed information regarding the study variables and their operational definitions are given in Table 1.

Table 1 *Operational Definitions of the Study Variables*

	Attribute	**Variable**	**Measurement type**	**Data type**	**Variable Definition**
1	Exogenous	Organizational Support	Latent		The extent organizational actors consider performance measurement (PM) as an important tool and therefore, support the use of performance measurement in their jurisdictions.
1.1	Exogenous - indicator	Support of Mayor	Measurable	Ordinal	The extent mayors consider PM as an important tool and therefore, support the use of performance measurement in their jurisdictions.
1.2	Exogenous - indicator	Support of Top Managers	Measurable	Ordinal	The extent top managers of a municipality consider PM as an important tool and therefore, support the use of PM in their jurisdictions.
1.3	Exogenous - indicator	Support of Middle Managers	Measurable	Ordinal	The extent middle managers of a municipality consider PM as an important tool and therefore, support the use of PM in their jurisdictions.
1.4	Exogenous - indicator	Support of Employees	Measurable	Ordinal	The extent employees of a municipality consider PM as an important tool and therefore, support the use of PM in their jurisdictions.
1.5	Exogenous - indicator	Special Meetings	Measurable	Ordinal	The frequency of special meetings held in the municipality to discuss performance measurement issues
1.6	Exogenous - indicator	Mayors' Communication about PM	Measurable	Ordinal	The frequency that mayors emphasize the importance of performance measurement in their communications with managers and other organizational actors
1.7	Exogenous - indicator	Internal Communication about PM	Measurable	Ordinal	The frequency that managers emphasize the importance of performance measurement in their communications with other managers and employees
2	Exogenous	External Support	Latent		The extent organizational actors consider PM as an important tool and therefore, support the use of performance measurement in their jurisdictions.
2.1	Exogenous - indicator	Support of Council Members	Measurable	Ordinal	The extent council members consider PM as an important tool and therefore, support the use of performance measurement in their jurisdictions.
2.2	Exogenous - indicator	Perceived Importance by Council Members	Measurable	Ordinal	The extent council members view performance measurement as an important aspect of decision making.
2.3	Exogenous - indicator	Council Meetings about PM	Measurable	Ordinal	The frequency of council meetings held to discuss performance measurement
2.4	Exogenous - indicator	Support of Citizens	Measurable	Ordinal	The extent citizens consider PM as an important tool and therefore, support the use of performance measurement in their jurisdictions.
2.5	Exogenous - indicator	Citizen Interest in Performance Data	Measurable	Ordinal	The extent citizens show their interest to the performance information the municipality provides.

	Attribute	Variable	Measurement type	Data type	Variable Definition
3	Endogenous – Mediating (M)	Technical Capacity	Latent		The extent the municipality has technical capacity to implement PM systems
3.1	Endog. (M) – Indicator	Staff	Measurable	Ordinal	The adequacy of staff number tasked with performance measurement
3.2	Endog. (M) – Indicator	Information Technology	Measurable	Ordinal	The adequacy of information technology and required equipments allocated for the implementation of PM systems
3.4	Endog. (M) – Indicator	Competency in Performance Measure Development	Measurable	Ordinal	The extent the staff can develop good performance measures
3.3	Endog. (M) – Indicator	Timely Collection of Performance Data	Measurable	Ordinal	The extent the staff can collect performance data in a timely manner
3.5	Endog. (M) – Indicator	Competency in Performance Data Analysis	Measurable	Ordinal	The extent the staff can assess and analyze the performance data
3.6	Endog. (M) – Indicator	Training	Measurable	Ordinal	Whether the staff attending regularly to conferences/workshops /trainings related to performance measurement
4	Endogenous – Mediating (M)	Quality of performance measures	Latent		To what extent PM systems are implemented in the municipality
4.1	Endog. (M) – First-order	Validity	Latent		The extent the performance measures are technically sound
4.1.1	Endog. (M) – Indicator	Derived from missions/goals	Measurable	Ordinal	The extent the performance measures are derived from missions and goals
4.1.2	Endog. (M) – Indicator	Derived from service standards	Measurable	Ordinal	The extent the performance measures are derived from service standards
4.1.3	Endog. (M) – Indicator	Focus on importance	Measurable	Ordinal	The extent the performance measures focus on what is important to measure, not on the availability of data
4.1.4	Endog. (M) – Indicator	Being up to date	Measurable	Ordinal	The extent the performance measures are current and up to date
4.1.5	Endog. (M) – Indicator	Being clear/understandable	Measurable	Ordinal	The extent the performance measures are clear and understandable

	Attribute	Variable	Measurement type	Data type	Variable Definition
4.1.6	Endog. (M) – Indicator	Measuring performance over time	Measurable	Ordinal	The extent the performance measures measure the performance over time
4.2	Endog. (M) – First-order	Legitimacy	Latent		The extent the performance measures are seen legitimate by the stakeholders
4.2.1	Endog. (M) – Indicator	Involvement of Managers	Measurable	Ordinal	The extent managers involve in the development process of performance measures
4.2.2	Endog. (M) – Indicator	Involvement of Employees	Measurable	Ordinal	The extent employees involve in the development process of performance measures
4.2.3	Endog. (M) – Indicator	Informing council members	Measurable	Ordinal	The extent the city council is informed about the efforts to develop performance measures.
4.2.4	Endog. (M) – Indicator	Perceived Usefulness by Elected Officials	Measurable	Ordinal	The extent the performance measures are perceived useful by elected officials
4.2.5	Endog. (M) – Indicator	Perceived Usefulness by Managers	Measurable	Ordinal	The extent the performance measures are perceived useful by managers
4.2.6	Endog. (M) – Indicator	Perceived Usefulness by Employees	Measurable	Ordinal	The extent the performance measures are perceived useful by employees
4.3	Endog. (M) – First-order	Functionality	Latent		The extent the performance measures have potential for creating benefits
4.3.1	Endog. (M) – Indicator	Potential for service quality improvement	Measurable	Ordinal	The extent the performance measures have potential for improving service quality
4.3.2	Endog. (M) – Indicator	Potential for decision-making capacity improvement	Measurable	Ordinal	The extent the performance measures have potential for improving decision-making capacity
4.3.3	Endog. (M) – Indicator	Potential for increasing employee motivation	Measurable	Ordinal	The extent the performance measures have potential for increasing employee motivation
4.3.4	Endog. (M) – Indicator	Potential for stimulating organizational learning	Measurable	Ordinal	The extent the performance measures have potential for stimulating organizational learning
4.3.5	Endog. (M) – Indicator	Potential for improving external communication	Measurable	Ordinal	The extent the performance measures have potential for improving external communication with elected officials and citizens

	Attribute	Variable	Measurement type	Data type	Variable Definition
5	Endogenous	Effectiveness of PM	Latent		To what extent PM systems in the municipality is effective
5.1	Endogenous - Indicator	Improvement in Productivity	Measurable	Ordinal	The extent the PM improves productivity in the municipality
5.2	Endogenous - Indicator	Improvement in Service Quality	Measurable	Ordinal	The extent the PM improves service quality in the municipality
5.3	Endogenous - Indicator	Increase in Employee Motivation	Measurable	Ordinal	The extent the PM increases the motivation of employees
5.4	Endogenous - Indicator	Stimulation of Organizational Learning	Measurable	Ordinal	The extent the PM stimulates organizational learning in the municipality
5.5	Endogenous - Indicator	Improved relationship with community	Measurable	Ordinal	The extent the PM improves the relations with the community.
5.6	Endogenous - Indicator	Cost Reduction	Measurable	Ordinal	The extent the PM creates reductions in the costs of our municipal services
5.7	Endogenous - Indicator	Better Identification of Problems	Measurable	Ordinal	The extent the PM helps the managers to better identify managerial and operational problems in municipal departments.
5.8	Endogenous - Indicator	Better Solution of Problems	Measurable	Ordinal	The extent the PM helps the managers to better develop solutions to managerial and operational problems in municipal departments
5.9	Endogenous - Indicator	Better Decision-making	Measurable	Ordinal	The extent the PM helps managers to make better decisions.
5.10	Endogenous - Indicator	Better Communication with Elected Officials	Measurable	Ordinal	This organization's performance measurement helps managers to communicate more effectively with elected officials.
6	Control	Population	Measurable	Ordinal	The population to which the municipality provide services
7	Control	Type of the Municipality	Measurable	Nominal	Whether a municipality is a metropolitan, metropolitan district, city, or district municipality

4.2 Data Collection

The data for the study were gathered by a cross-sectional survey sent to all 1000 (Turkish Ministry of Interior, 2014b) Turkish municipalities, which are province and district municipalities including metropolitan municipalities. The reason of selecting this criterion is based on the assumption that the utilization of performance measurement systems is low in smaller localities (Folz et al., 2009; Riverbank & Kelly, 2003) and town municipalities do not have qualified personnel experienced in performance measurement and as a result, their responses may be given by people who have very limited knowledge about performance measurement. The surveys were completed by senior officials of the municipalities.

Although using a survey as data collection method may create some limitations, most of the empirical studies (such as Ammons & Rivenbark, 2008; Berman & Wang, 2000; de Lancer Julnes & Holzer, 2001; Ho, 2006; Ingraham et al. 2003, Streib & Poister, 1999; Poister & Streib, 1999; Taylor, 2006, 2011; Wang & Berman, 2001, Yang & Hsieh, 2007) in the area of performance measurement used this method. According to Dillman, Smyth, and Christian (2009), surveys can create generalizable results, if they are prepared and implemented correctly.

The questions in the survey regarding the quality of performance measures are directly taken from Streib and Poister (1999). The other questions are directly taken or adapted from several studies (Berman & Wang, 2000; de Lancer Julnes & Holzer, 2001; Yang & Hsieh, 2007). All question groups in the survey use a five-point Likert scale from strongly disagree (coded 1) to strongly agree (coded 5). The survey has a total of 55 questions, three of which are open-ended questions. The survey was expected to be completed between 10-15 minutes.

The survey was conducted in Turkish. In order to ensure the validity of the translation, it was reviewed by a group of Turkish scholars and PhD students. The survey questionnaire and its Turkish version are given in Appendices A and B respectively. Following the revisions, the survey was sent to municipalities via e-mail, based on an e-mail list taken from Turkish Ministry of Interior. Qualtrics, which is an online survey tool, was used to manage the processes of survey distribution and data collection. The reason for the utilization of an online survey tool was its advantages regarding low cost and easy access to many respondents, and ease in making the data ready for the analysis.

In order to increase the response rate of the survey, a document showing the support of the Turkish Ministry of Interior for the study was posted to the municipalities. Moreover, in order to increase the response rate, the survey was sent to the municipalities four times following the recommendations of Dillman et al. (2009). Utilization of the online survey tool made it much easier that the follow-up e-mails were sent to those who did not complete the survey yet.

In the study, confidentiality of survey responses was preferred over the anonymity of them. Although, anonymity could contribute to the validity of the responses by decreasing the risk that the respondents give socially desirable responses, it is also possible that the fact that nobody will observe the response could lead the busy respondents to pay little attention to their responses. Moreover, anonymity would create significant difficulties in following the municipalities that responded the survey, which is important for indicating whom to send the later waves of survey e-mails.

4.3 Statistical Analysis

In the study, several statistical analyses were used to analyze the data collected from the survey. These analyses are descriptive analysis including correlation analysis, confirmatory factor analysis, reliability analysis, and structural equation modeling. A brief explanation of how these analyses were utilized in this study is explained in this section.

4.3.1 Descriptive Analysis

In order to examine the main characteristics of the data, firstly, descriptive statistics was run by using the SPSS program. Descriptive statistics are used to analyze the distributional characteristics of the data. In this study, frequency tables of each study variable are given separately to illustrate how the responses to survey questions are distributed. Descriptive statistics are also useful to detect the presence of any missing data.

Another important function of the descriptive analysis is to examine the correlations among the indicators of the latent variables and to detect if multicollinearity exists among them. In the study, Spearman's rank order correlation, which is mostly known as Spearman's rho, is used to examine the correlations among the observed variables, since the observed variables in the study are ordinal variables and their correlation is better examined by this method (Kline, 2011). Multicollinearity occurs if two indicators of a latent variable are highly correlated to each other, which means that they mainly measure the same thing (Kline, 2011). As a result, using both of these indicators is redundant and one of them should be removed from the model. The scholars mention several thresholds for deciding the presence of multicollinearity among the variables. Scholars like Meyers, Gamst, and Guarino (2013) consider .70 as the threshold for detecting high correlation, which may be a sign of multicollinearity. On the other hand, some scholars like Kline (2011) and Garson (2012) put a

higher threshold, .85, for multicollinearity. In this study, .85 was set as the threshold for detecting multicollinearity.

4.3.2 Confirmatory Factor Analysis

Following the descriptive analysis, confirmatory factor analysis (CFA) was conducted by using AMOS 22 software to evaluate and validate the measurement models of the latent study variables. Since the latent constructs cannot be measured directly, measurement models composed of several indicators are utilized to measure these constructs (Bryne, 2010). However, it is important to check the validity of these measurement models before proceeding to the SEM analysis. This is done by using confirmatory factor analysis (Wan, 2002; Bryne, 2010). According to Byrne (2010), factor analysis is "the oldest and best-known statistical procedure for investigating relations between sets of observed and latent variables" (p. 5). As explained by Wan (2002), factor analysis aims "to simplify complicated and diverse relationships among variables by revealing common factors that link seemingly unrelated variables" (p. 55). Confirmatory factor analysis is a type of factor analysis, which is "designed to test the hypothesized link between the observed variables and known underlying factors" (Wan, 2002, p. 55).

In the study, measurement models will be evaluated and, if needed, revised by using a three-step method proposed by Wan (2002). These steps are; 1) checking the appropriateness of the indicators, 2) checking the overall model fit, 3) revising the model. These steps will be explained in more depth in the later sections.

Since there are five latent variables in the study, five measurement models, which show how these latent constructs will be measured, are needed. The proposed measurement models are presented in the following section.

4.3.2.1 Measurement model for organizational support. Organizational support was measured in the study with the extent of support from the mayor, top managers, lower level managers, and employees for the use of performance measurement in their municipality. In addition to direct questions regarding the support of these stakeholders, the study also used some indirect questions to understand the level of organizational support. These questions are regarding the frequency of the special meetings held in which performance measurement is discussed, the frequency of mayor's communication about the importance of performance measurement to organizational actors, and the frequency of top-down internal (from managers to employees) communication about the issues related to performance measurement (adapted from de Lancer Julnes and Holzer, 2001). Proposed measurement model for organizational support is presented in Figure 5.

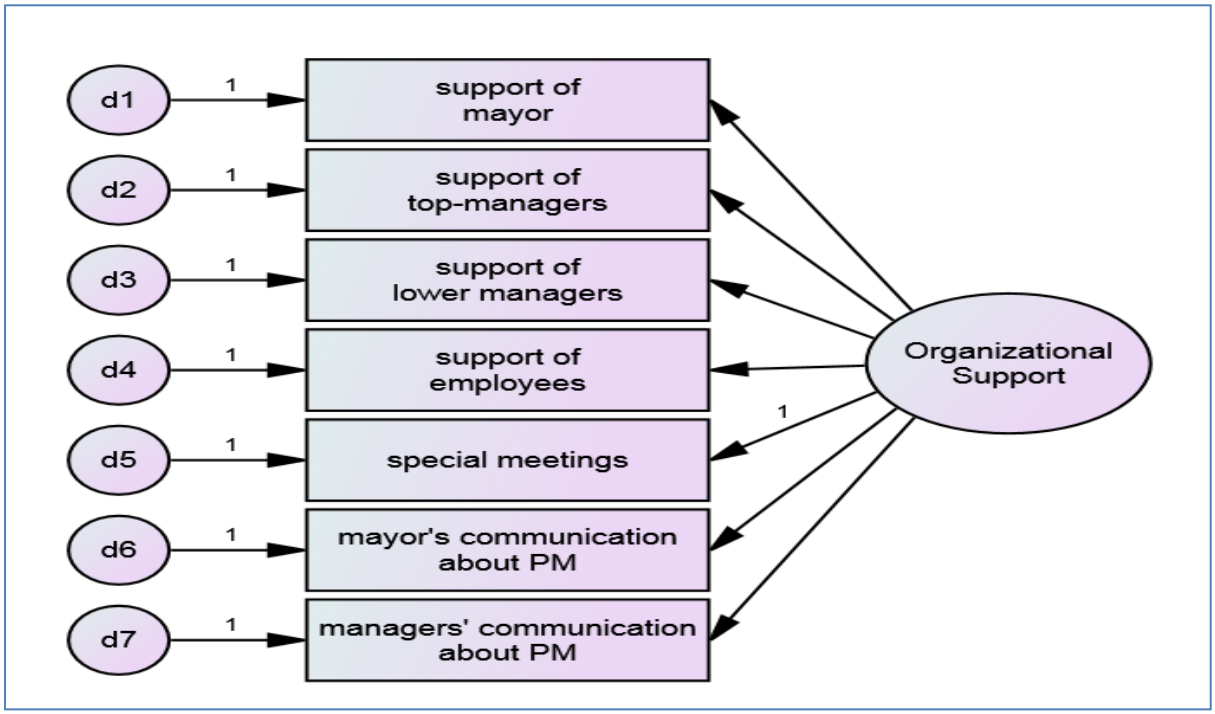

Figure 5. Proposed Measurement Model for Organizational Support

4.3.2.2 Measurement model for external support. External support was measured in the study with the extent of support from two groups of stakeholders: council members and citizens. Similar to organizational support, the survey used indirect questions to understand the level of external support, in addition to direct questions regarding the support of these stakeholders. These indirect questions are regarding the extent the council members view performance measurement as an important aspect of decision making, the frequency of the council meetings in which performance measurement or data is discussed, and the extent citizens show their interest to municipality's performance information. Proposed measurement model for external support is presented in Figure 6.

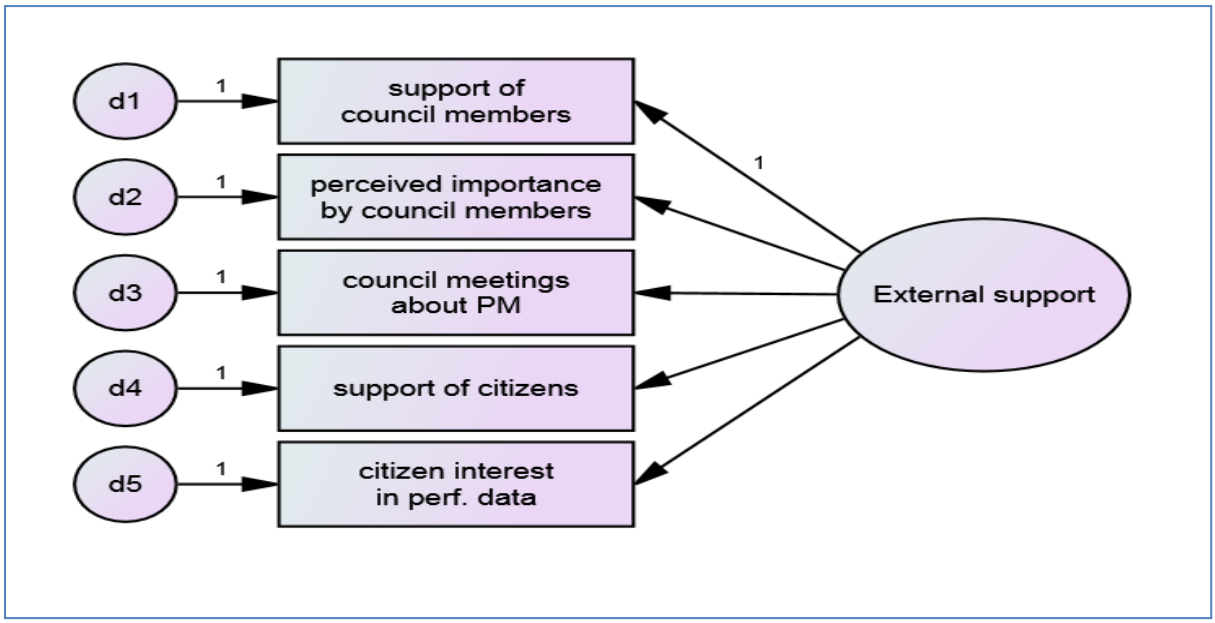

Figure 6. Proposed Measurement Model for External Support

4.3.2.3 Measurement model for technical capacity. In the study, technical capacity has six indicators, which are mostly adapted from Berman and Wang (2000) and Yang and Hsieh (2007). These indicators are the adequacy of the staff and information technology, the competencies of the staff for the development of high quality performance measures, timely collection of performance data, and the analysis of performance data, and the frequency the

staff attends to training activities. Proposed measurement model for technical capacity is presented in Figure 7.

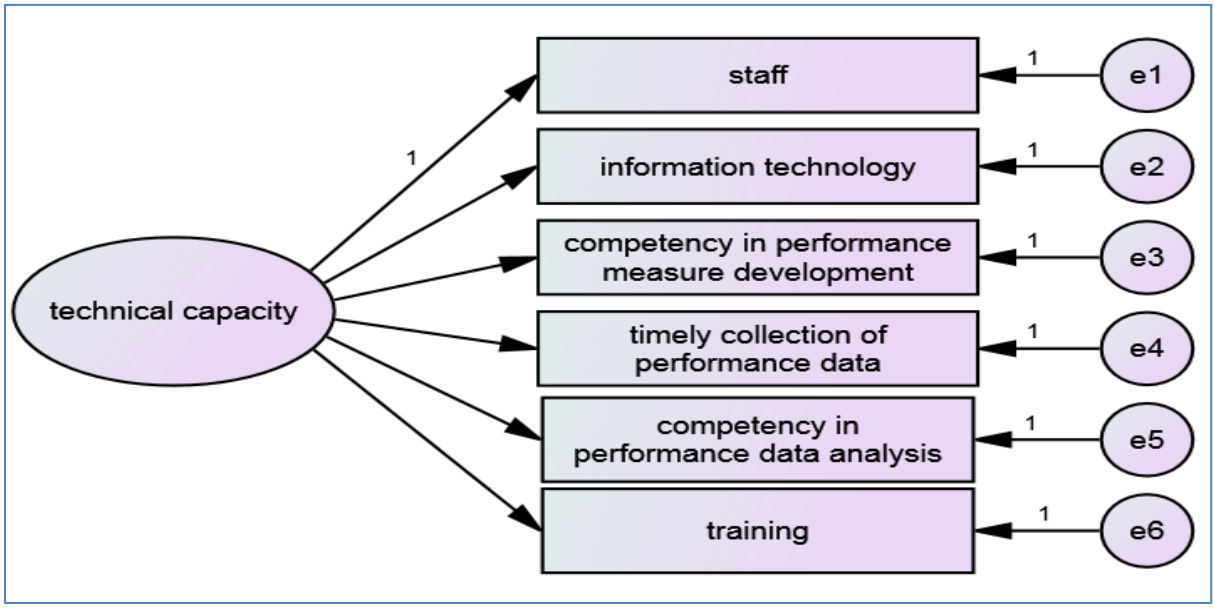

Figure 7. Proposed Measurement Model for Technical Capacity

4.3.2.4 Measurement model for quality of performance measures. Quality of performance measures is related to the adoption stage of performance measurement. As mentioned earlier, this study used Bouckaert's (1993) model in assessing the quality of performance measures. According to this model, performance measures can be considered as effective, if they are valid, legitimate, and functional. As a result, measurement model for quality of performance measures includes these three dimensions. The indicators in the study regarding these dimensions are adapted from Streib and Poister (1999). In order to measure validity, the study seeks answers about the extent that measures are developed from organizational missions, goals, and service standards, the extent they focus on what is important to measure (not the availability of data), the extent they are up to date and clear, and the extent they track performance over time. For legitimacy, it is important to find out that to what extent managers and lower level employees involve in the development of

performance measures, and the extent they and elected officials perceive developed performance measures useful. The functionality dimension is more related to benefit creating potential of performance measures and aims to find out the extent that developed performance measures have the potential to improve service quality, decision-making capacity, employee motivation, organizational learning, and communication of managers with elected officials. Proposed measurement model for the quality of performance measures is presented in Figure 8.

4.3.2.5 Measurement model for effectiveness of performance measurement.

Effectiveness of performance measurement is related to the implementation phase of performance measurement and concerns mainly the effects of performance measurement (Yang & Hsieh, 2007). The indicators of this measurement model are mainly adapted from Yang and Hsieh (2007). Several effects of performance measurement are pointed out in the literature. This study examined to what extent performance measurement improves productivity and service quality, increases employee motivation, stimulates organizational learning, improves relations with the community, helps managers to identify the problems and create solutions for these problems, facilitates better decisions, and contributes to the managerial communication with elected officials. Proposed measurement model for the effectiveness of performance measurement is presented in Figure 9.

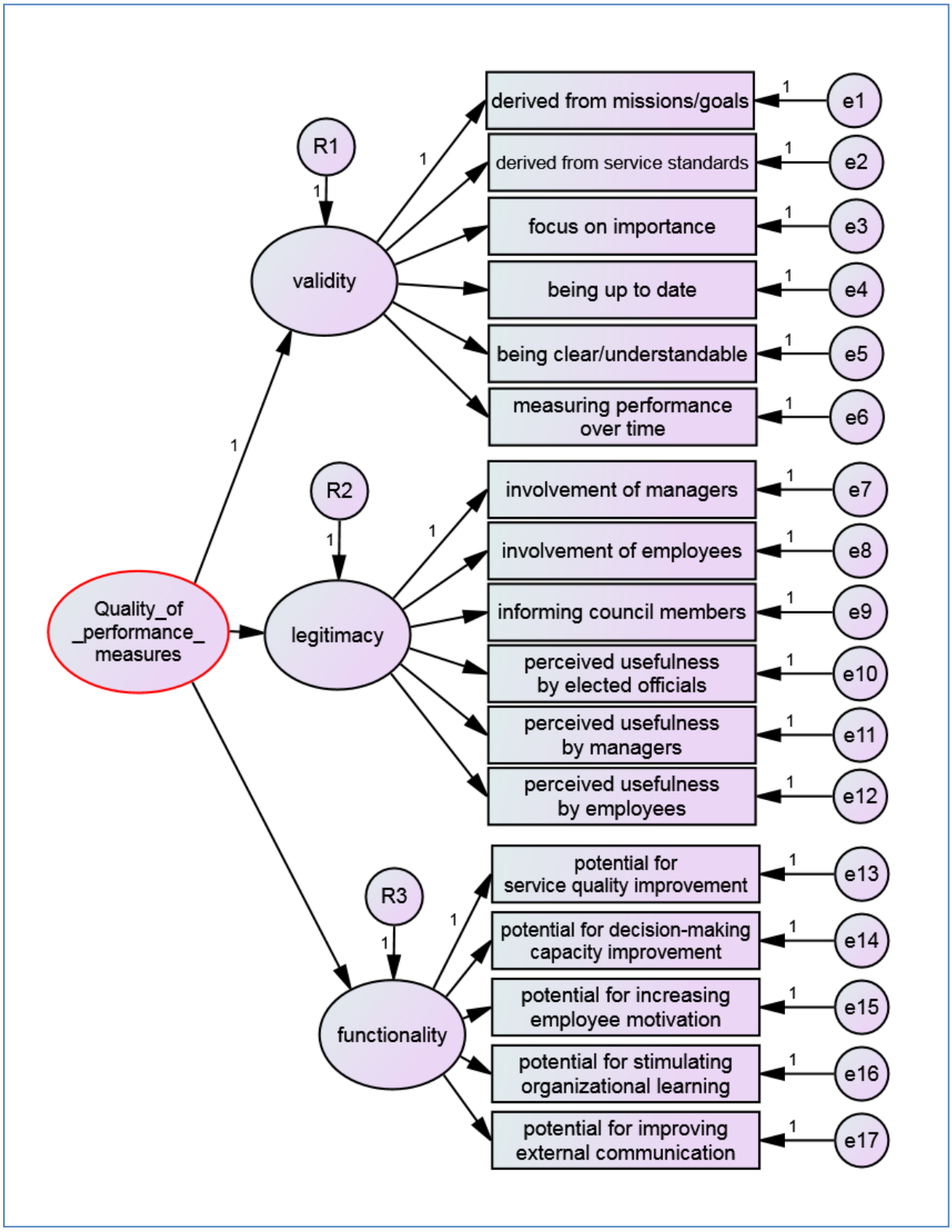

Figure 8. Proposed Measurement Model for Quality of Performance Measures

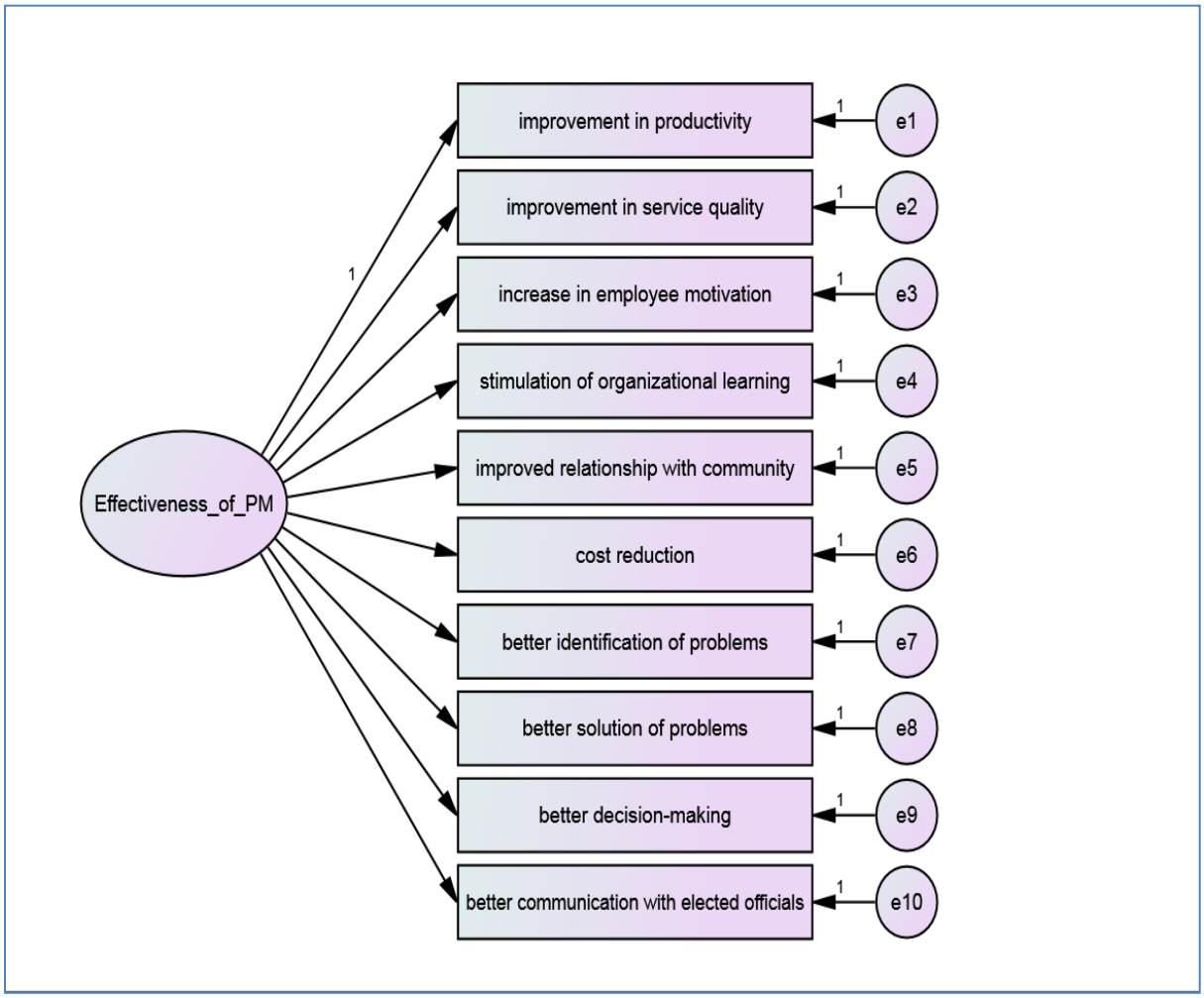

Figure 9. Proposed Measurement Model for Effectiveness of Performance Measurement

4.3.3 Reliability Analysis

Measurement reliability is an important part of any study which includes scale variables. Measurement reliability concerns on non-systematic or random errors. It mainly examines whether measurement create consistent results over time (Babbie, 2013). Since the scales for latent variables in the study are not taken from the literature as a whole, it is also important to test the measurement reliability of the scales used. One of the most common methods in measuring the reliability is using Cronbach's alpha score (Kline, 2011; Streib & Poister, 1999). If this score exceeds the generally accepted adequate level of .70 (de Lancer Julnes & Holzer, 2001; George & Mallery, 2007; Kline, 2011; Morgan, Leech, Gloekner, &

Barrett, 2005), it means that the measurement produces consistent results at different times (Cronbach, 1951). An alpha score greater than .80 is considered as good and greater than .90 is considered as excellent (George & Mallery, 2007). In this study, .70 was set as the threshold for Cronbach's alpha. If the alpha score does not meet this criterion, then the measurement scales need to be revised.

4.3.4 Structural Equation Modeling

In the study, SEM was used to examine the hypothesized relationships between the study variables. SEM is a method which is useful for the analysis of causal links among variables in a combined structure model (Wan, 2002). As explained by Benson and Hagtvet (1996), "SEM is a general data analytic technique that subsumes many statistical … procedures [such as] analysis of variance and covariance, correlation, regression, factor analysis, and reliability estimation" (as cited in Yang, 2002, p. 305).

Yang and Hsieh (2007) compare SEM with other multivariate techniques and argue that it "has a stronger ability to test mediating relationships, test models with multiple endogenous variables, test overall models rather than individual coefficients, use confirmatory factor analysis to reduce measurement error, and take into account error terms" (p. 866). Since this study has multiple latent variables, multiple endogenous variables, and multiple mediating variables, SEM has been preferred as the appropriate statistical method to examine the relationships in the complex conceptual model proposed in the study.

Based on the explanations above and the hypotheses of the study, the covariance structure model of the study, which shows both the measurement models and the relationships between variables, is presented in Figure 10.

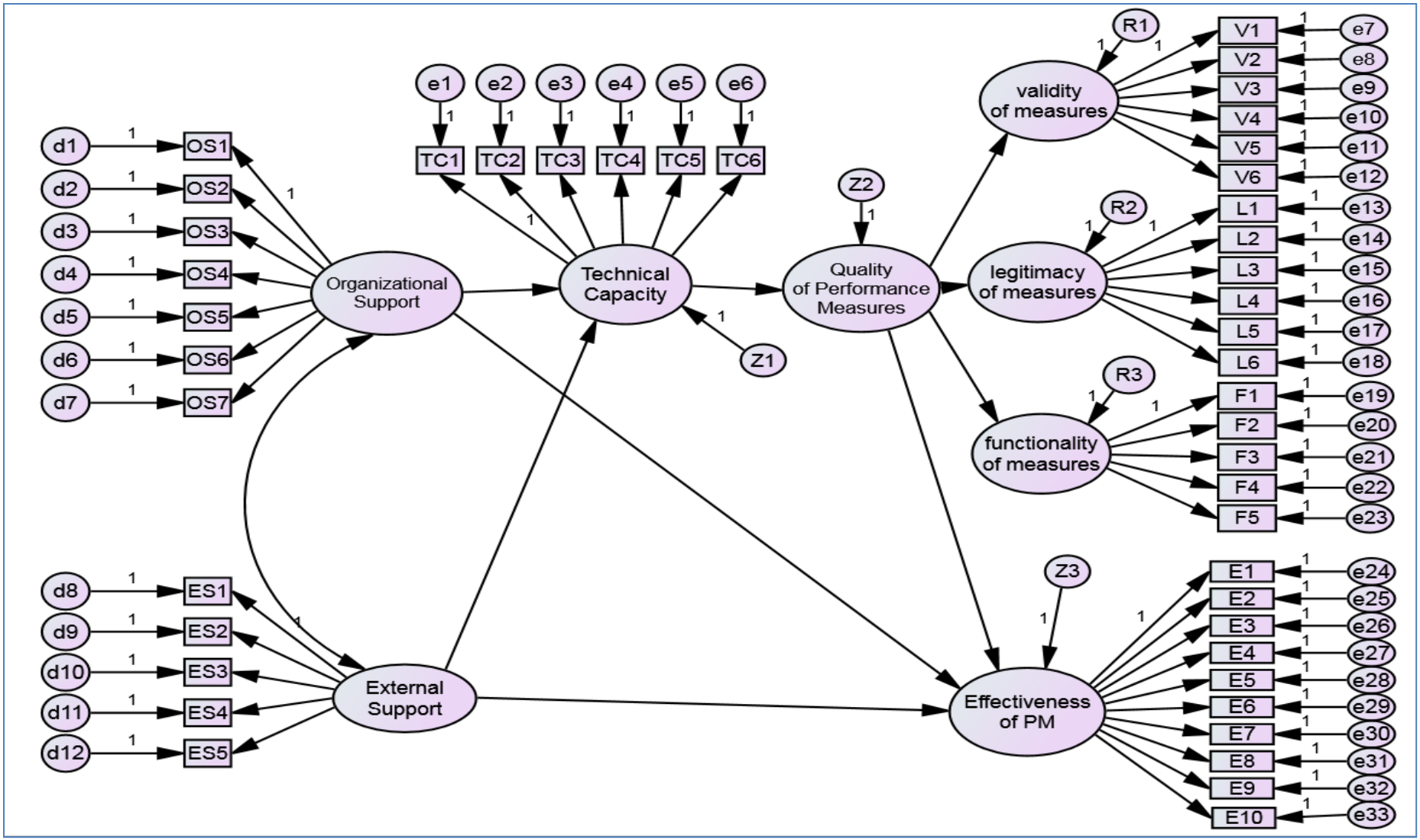

Figure 10. Proposed Covariance Structure Model

4.4 Power Analysis and Sample Size

In order to get meaningful results from the statistical analysis it is essential to have adequate power and sample size. Power analysis indicates the probability of rejecting a null hypothesis when it is false (Kaplan, 1995). It is a judgment call of the researcher about the desired precision of the results. Power of a study depends on the alpha level and sample size of the study. In this study, an alpha level of .05, which is the mostly used level in social sciences, will be used indicating that the results will be 95% confident and are not found by chance.

It should be noted that SEM is a "large sample technique", which means that "model estimation and statistical inference or hypothesis testing regarding the specified model and individual parameters are appropriate only if sample size is not too small" (Lei & Wu, 2007, p. 36; Ullman, 2006). There are several methods for identifying the required minimum sample size for a SEM model. In the literature, a minimum of 200 cases is suggested for most models (Boomsma & Hoogland, 2001; Kline, 2011). However, complexity of the model may necessitate a larger sample. A rule of thumb suggested in the literature for the optimal number of sample size is that the number of cases should be 5 to 20 times of the number of parameters (Bentler & Chou, 1987; Kline, 2011; Lei & Wu, 2007). However, it is also important to note that "it is possible to have results that are highly significant (e.g., $p < .0001$) but trivial in absolute magnitude when the sample size is large" (Kline, 2011, p. 13). In order to avoid an excessive sample, the optimal sample size for the study is calculated by multiplying the number of parameters with five. Since there are 106 parameters in the model, the optimal sample size for the study is calculated as 530. Consequently, the targeted number of responses in the study was between 200 and 530. The study has 428 samples, indicating that the sample size of the study is at the acceptable level.

4.5 Validation of the Models

The validation of the proposed model took place in two stages. In the first stage, each measurement model was evaluated and, if needed, revised by using a three-step method proposed by Wan (2002). In the second stage, the validated measurement models was integrated into the covariance structure model (CSM) and the CSM was evaluated and revised by using the goodness of fit statistics and modification indices.

Three steps used to evaluate and revise the measurement models are; 1) checking the appropriateness of the indicators, 2) checking the overall model fit, 3) revising the model. In this section these steps will be explained in detail.

The first step is about checking the appropriateness of the indicators. In this step, firstly, the presence of multicollinearity was examined. If multicollinearity among two indicators existed, one of the indicators was removed from the model. Secondly, the critical ratios and p values of regression weights for each indicator were examined and the indicators which have a critical ratio lower than 1.96 or higher than -1.96 and p values higher than .05 were excluded from the model, since these ratios show that there is not a significant relationship between the indicator and its latent construct at a .05 confidence level (Byrne, 2010). Thirdly, strengths of standardized regression weights of indicators were examined and the indicators which show a value lower than .30 were removed from the model, since factor loadings are only meaningful if they are greater than .30 (Hoe, 2008).

In the second step, overall model fit was checked by using the goodness of fit statistics generated by the AMOS software. Goodness of fits statistics show that how well the study model fits the actual data collected. A detailed discussion of which goodness of fit

statistics were used in this study and which criteria were selected as the cut off points of these statistics are given in later sections.

If the goodness of fit statistics does not show a good model fit, the search for the possible reasons of the lack of fit takes places, which constitutes the third step of Wan's (2002) method. In this step, examination of the modification indices is important in order to figure out which correlated errors should be freely estimated in order to reduce the chi-square value and fit the model better. Beginning from the highest modification indices, correlations between several measurement errors are identified and nested measurement models for the study variables are built. Another method for the modification of the models is the exclusion of some indicators from the model. When the revisions with modification indices did not create acceptable model fit, some of the indicators were excluded from the measurement models.

After the validation of measurement models, the next step is to validate the covariance structure model. The first step in this part is to examine whether 1) gamma effects (path coefficients) between the study variables, including control variables, 2) factor loadings, and 3) correlations among measurement errors are statistically significant. Insignificant relationships should be excluded from the model. Moreover, like measurement models, goodness of fit statistics should be used for validating the covariance structure model. If these statistics meet the criteria, mentioned below, this means that the model fits the data well and it is validated.

4.6 Overall Model Fit and Goodness of Fit Statistics

Goodness of fit statistics is useful to determine the extent to which the study model fits the data used for the analysis. Having an ability to test the model's fit is considered as one

of the important advantages of SEM (Schumacker & Lomax, 2010). Based on the results of the goodness of fit indices, 1) the model can be accepted, 2) a need for improvement may arise, or 3) the model may be required to be rejected. If the goodness of fit statistics don't meet the threshold criteria, this means that path coefficients or regression weights in the model don't have any meaning (Garson, 2012).

There are several goodness of fit statistics which are used to test the fit of the model. For example, AMOS produces 25 different goodness of fit measures (Garson, 2012). In the literature, there is not any agreement on which statistics to be reported and it is recommended that a group of indexes should be reported when accepting or rejecting a model fit (Byrne, 2010; Garson, 2012; Schumacker & Lomax, 2010).

In their literature review about the published SEM papers, McDonald and Ho (2002) found that the papers mostly reported comperative fit index (CFI), goodness of fit index (GFI), normed fit index (NFI) and nonnormed fit index (NNFI or also known as Tucker-Lewis index – TLI). However, Garson (2012) argues that GFI and adjusted goodness of fit index (AGFI) are not recommended, since they underestimate the fit of complex models and they are sensitive to sample size. Kline (2011) recommends the use of root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), CFI, model chi-square (CMIN), its degrees of freedom and associated p value. Similarly, Garson (2012) put special emphasis on reporting model chi-square, RMSEA, and one of the following measures: incremental fit index (IFI), relative fit index (RFI), NFI, TLI, or CFI. Thompson (2000) argued that RMSEA and CFI are the most useful indices in the assessment of a model fit. Garver and Mentzer (1999) make a similar recommendation by including TLI in addition to RMSEA and CFI. Also recommended in the literature (Garson, 2012) is the Hoelter Index (also known as Hoelter's Critical N), which is useful to judge whether the sample size is

adequate for the model. In this study, model chi-square (CMIN or χ^2), its associated p value, relative chi-square (CMIN/df), RMSEA, TLI (NNFI), CFI, and Hoelter's Critical N were used to determine the fit of the study models.

The CMIN (χ^2) is one of the widely utilized indicators of model fit (Garson, 2012). A CMIN value equals to zero indicates a perfect fit and smaller values mean a better fit. A statistically significant p value of CMIN shows that there is a significant discrepancy between the proposed study model and the actual model, meaning a poor model fit. In that sense, researchers aim p values to be higher than .05. However, CMIN is criticized by its sensitivity to sample size (Garson, 2012). In a larger sample size, it is more likely to flag even very trivial differences as significant and to reject something true (type II error) (Ullman, 2007). Therefore, Garson (2012) and Schermelleh-Engel, Moosbrugger, and Müller (2003) argue that for many researchers finding significant CMIN is not a reason for model modification, if sample size is over 200 and other fit indices show a good fit. In the literature, relative chi-square (CMIN/df) test, which is an attempt to make CMIN less dependent on sample size, is recommended (Garson, 2012). There are several suggestions in the literature regarding the cut off value for the relative chi-square, ranging from 2 (Ullman, 2007) to 5 (Schumacker & Lomax, 2010). In this study, the cut-off value for CMIN/df was determined as 4, meaning the values lower than 4 will be considered as a good model fit.

According to Schermelleh-Engel et al. (2003), reporting RMSEA, TLI, and CFI should always be considered, since these measures are sensitive to model misspecifications. RMSEA is measure of approximate fit and "concerned with the discrepancy due to approximation" (Schermelleh-Engel et al., 2003, p. 36). It mainly examines the extent the proposed model is close to the actual model. RMSEA takes degrees of freedoms into account and is sensitive to the complexity of the model (Bryne, 2010). Several cutoff values are

recommended in the literature such as .05 (Schumacker & Lomax, 2010, Wan, 2002), .06 (Hu & Bentler, 1999), .08 (Garson, 2012, Sivo et al., 2006, Wan, 2002), and .1 (Bryne, 2010). In this study, RMSEA values lower than .08 are considered as a good model fit.

TLI, also known as NNFI, is another measure of approximate fit. It takes degrees of freedoms into account when comparing null model fit with the study model and therefore, it is argued that it is not sensitive to sample size. Simpler models are rewarded in TLI. As a result, it is highly recommended for the evaluation of models (Garson, 2012). Similarly recommended in the literature is the CFI, also known as Bentler comparative fit index, which is developed to avoid the problem of underestimation of fit by NFI in small sample sizes (Bryne, 2010; Schermelleh-Engel et al., 2003). CFI is a desirable index for the comparison of nested models (Byrne, 2010; Ullman, 2007). The values of TLI and CFI range from 0 to 1, where 1 indicates the perfect fit. For these indices, the cut-off points of .90 (Garson, 2012) and .95 (Byrne, 2010; Hu & Bentler, 1999) are recommended for the model fit. In the study .90 was taken as the threshold level for TLI and CFI.

Lastly, Hoelter Index (also known as Hoelter's Critical N) will be examined in the evaluation of model fit. This index will help to determine the extent that the study has adequate sample size for the evaluation of model fit. The values equal to or higher than 200 are considered adequate sample size, and the values between 75 and 200 are considered as acceptable sample size (Garson, 2012). The criteria, taken in the study, for indicating that the study models fit the data at hand are presented in Table 2.

Table 2 *Goodness of Fit Indices*

Fit Index	Criteria
Chi-Square (CMIN)	Smaller the better
Chi-Square related p value	≥ .05
Chi-square / Degree of Freedom (CMIN /df)	≤ 4
Tucker Lewis Index (TLI)	≥ .90
Comparative Fit Index (CFI)	≥ .90
Root Mean Square Error of Approximation (RMSEA)	≤ .08
Hoelter's Critical N	≥ 75

4.7 Human Subjects

Since the study collected data from human subjects, it is obligatory to take the approval of the UCF Institutional Review Board (IRB) before implementing the survey. Therefore, the approval was taken and it is presented in Appendix C. Moreover, the respondents were informed that the participation to this survey is voluntary and they can quit the survey any time they want. Furthermore, all responses to the survey questions were kept confidential, and will not be revealed without consent of the respondents; only aggregate results will be made available.

In this chapter, study variables, their measurement models, the methods of data collection and data analysis were explained in detail. In the next chapter, findings of these statistical analyses will be presented.

CHAPTER 5: FINDINGS

This chapter presents the results found by the statistical analyses explained in the previous chapter. Firstly, results of descriptive analysis of the variables will be given for each study variable, in order the reader to get a sense of the frequency distributions of the survey responses. Secondly, the findings of correlation analysis will be explained. If high correlation among the indicators of a latent construct is found, one of the indicators will be removed from the model. Then, with the confirmatory factor analysis the measurement models will be validated and insignificant or unimportant indicators will be removed from the models. The validated models will then be examined through the reliability analysis. The scales which show low reliability will be revised. After completing each of these steps, all measurement models and control variables will be combined in a covariance structure model and this model will be analyzed with structural equation modeling (SEM). Based on the results of the SEM, whether the hypotheses of the study are supported or not supported will be explained in the last part of the chapter.

5.1 Descriptive Analysis

In this section of the study, descriptive statics, which were provided by the SPSS program, will be presented. As mentioned, descriptive statistics are used to analyze the distributional characteristics of the data. Frequency tables of each study variable are given below to illustrate how the responses to survey questions are distributed.

The survey was sent to all 1,000 (Turkish Ministry of Interior, 2014b) Turkish municipalities, which are province and district municipalities including metropolitan municipalities. A total of 580 municipalities responded to the survey, however, 152 of these responses had missing data. Since the responses with missing data have considerable missing

data, they were excluded from the data set. A total of 428 complete responses were used in the data analysis. Consequently, the response rate for the survey is 42.8%.

The surveys were expected to be completed by officials, either in managerial or expert level, who have a deep knowledge about the performance measurement practices of the municipality. The surveys were responded mostly by lower level managers and the experts with a total of 251 and 113 respectively. Moreover, 10 mayors, 16 deputy mayors, and 38 other top-managers responded the survey. In general, those who responded the survey are experienced officials with around 53% working in the municipality more than 10 years and another 26% working between 3 and 10 years. Only 8% was working less than 1 year. However, only 35% of the respondents have been working in the performance measurement practices more than 3 years. Yet, it is not an unexpected result considering the relatively new meeting of the Turkish municipalities in the last decade with the concept of performance measurement. Moreover, local elections took place in 31st of March, 2014. It is customary in Turkey that some replacements are done in municipal positions if the mayor of municipality changes.

Since not all the municipalities responded to the survey, it is important to evaluate the extent that the sample municipalities which responded to the survey can represent the all municipality population. Therefore, a comparison between the respondent municipalities and all municipalities is needed. In the study this comparison is done in two aspects: type and population of the municipalities, which are control variables of the study.

The distribution of the responses according to the types of the municipalities is almost proportionate to the actual distribution of municipalities in Turkey. The comparison of the sample in the population according to the type of the municipality is presented in Table 3.

Table 3 *Frequency and Percentage Distributions for the Type of Municipality*

Type	Response Frequency	Actual Frequency	Response Percentage	Actual Percentage	Difference Percentage
Metropolitan	17	30	4	3	1
Metropolitan District	196	519	45.8	51.9	-6.1
City	30	51	7.0	5.1	1.9
District	185	400	43.2	40	3.2
Total	428	1.000	100	100	

In Turkey, there are 30 metropolitan, 51 city, 519 metropolitan district, 400 district municipalities, which equal to 3%, 5,1%, 51,9%, and 40% of all municipalities respectively. The distribution of responses is similar to these percentages. Only metropolitan districts seem to be underrepresented in the sample with a 6.1%. The probable reason for this underrepresentation is the formation of 14 metropolitan and 26 metropolitan district municipalities. Metropolitan municipalities were formed by transforming the task and responsibility area of existing city municipalities. As a result, they inherited also performance measurement practices and experience of the previous municipalities. On the other hand, most of the metropolitan district municipalities were newly founded and they needed time to form their structure and processes. Therefore, they may have very little, if any, to say about their performance activities. As a result, it can be considered as normal that metropolitan district municipalities are somewhat underrepresented.

On the other hand, the comparison of respondents to all municipalities according to population of municipalities reveals an increasing trend as the population of municipality increases. The comparison table is given in Table 4.

Table 4 *Frequency and Percentage Distributions for Population*

Population	Response Frequency	Actual Frequency	Response Percentage	Actual Percentage	Difference Percentage
Less than 10.000	114	322	26.6	32.2	-5.6
Between 10,000 and 50,000	125	339	29.2	33.9	-4.8
Between 50,001 and 10000	59	111	13.8	11.1	2.7
Between 10001 and 250,000	64	118	15.0	11.8	3.2
More than 250,000	66	110	15.4	11	4.4
Total	428	1000	100	100	

The table shows that municipalities with higher populations are more represented in the sample. However, this is an expected finding based on the literature saying that cities with higher population utilize performance measurement more than those with lower populations (Folz et al. 2009; Riverbank & Kelly, 2003). As a result, it can reasonably be expected that they have more experience to share about performance measurement.

5.1.1 Organizational Support

Organizational support, which is an exogenous variable in the model, concerns the level of support from organizational actors, such as the mayor, top managers, lower level managers, and employees, for the use of performance measurement in the municipality. In the survey, respondents were asked seven questions to understand the level of organizational support in their municipality. The frequency table for the indicators of organizational support is given in the Table 5.

Table 5 *Frequency and Percentage Distributions for Organizational Support*

Indicator	Response	Frequency	Percentage	Cumulative Percentage
Support of Mayor (OS1)	Strongly Disagree	4	.9	.9
	Disagree	15	3.5	4.4
	Neither Agree nor Disagree	51	11.9	16.4
	Agree	188	43.9	60.3
	Strongly Agree	170	39.7	100
	Total	428	100	
Support of Top-managers (OS2)	Strongly Disagree	11	2.6	2.6
	Disagree	22	5.1	7.7
	Neither Agree nor Disagree	60	14.0	21.7
	Agree	192	44.9	66.6
	Strongly Agree	143	33.4	100
	Total	428	100	
Support of Lower Level Managers (OS3)	Strongly Disagree	10	2.3	2.3
	Disagree	23	5.4	7.7
	Neither Agree nor Disagree	50	11.7	19.4
	Agree	221	51.6	71.0
	Strongly Agree	124	29.0	100
	Total	428	100	
Support of Employees (OS4)	Strongly Disagree	15	3.5	3.5
	Disagree	44	10.3	13.8
	Neither Agree nor Disagree	97	22.7	36.4
	Agree	192	44.9	81.3
	Strongly Agree	80	18.7	100
	Total	428	100	
Special Meetings (OS5)	Strongly Disagree	30	7.0	7.0
	Disagree	92	21.5	28.5
	Neither Agree nor Disagree	79	18.5	47.0
	Agree	170	39.7	86.7
	Strongly Agree	57	13.3	100
	Total	428	100	
Mayor's Communication about PM (OS6)	Strongly Disagree	23	5.4	5.4
	Disagree	63	14.7	20.1
	Neither Agree nor Disagree	75	17.5	37.6
	Agree	169	39.5	77.1
	Strongly Agree	98	22.9	100
	Total	428	100	
Managers' Communication about PM (OS7)	Strongly Disagree	25	5.8	5.8
	Disagree	68	15.9	21.7
	Neither Agree nor Disagree	96	22.4	44.2
	Agree	181	42.3	86.4
	Strongly Agree	58	13.6	100
	Total	428	100	

The first four questions were about the extent of support from mayor, top managers, lower level managers, and employees for the use of performance measurement. The respondents agreed or strongly agreed that mayors, top and lower level managers have a considerable support for the use of performance measurement with 83.6%, 78.3%, 80.6%, respectively. However, the responses show an important difference with the level of support from non-managerial employees compared to these actors. Only 63.6% of the employees seem to support performance measurement practices according to the results.

In addition to direct questions regarding the support of these stakeholders, the study also used some indirect questions to understand the level of organizational support. These questions are regarding the frequency of the special meetings held in which performance measurement is discussed, the frequency of mayor's communication about the importance of performance measurement to organizational actors, and the frequency of top-down internal (from managers to employees) communication about the issues related to performance. More than half of the respondents agreed or strongly agreed that these meetings related with performance measurement or performance data are frequently held in their municipalities. On the other hand, 28% reported the absence of these meetings. Regarding the communications of mayor and managers, 62.4% of the respondents reported that their mayor frequently emphasizes the importance of performance measurement, whereas only 55,9% reported such a frequent emphasis by the managers.

In total, most of the respondents, ranging from 55% to 83% for different indicators, reported their agreement on the positive statements regarding the indicators of organizational support. These results show that there is considerable organizational support for the use of performance measurement in Turkish municipalities.

5.1.2 External Support

External support, which is also an exogenous variable in the model, concerns the level of support from external actors, such as council members and citizens, for the use of performance measurement in the municipality. In the survey, respondents were asked five questions to understand the level of external support in their municipality. The frequency table for the indicators of external support is given in the Table 6.

Table 6 *Frequency and Percentage Distributions for External Support*

Indicator	Response	Frequency	Percentage	Cumulative Percentage
Support of Council Members (ES1)	Strongly Disagree	25	5.8	5.8
	Disagree	50	11.7	17.5
	Neither Agree nor Disagree	109	25.5	43.0
	Agree	187	43.7	86.7
	Strongly Agree	57	13.3	100
	Total	428	100	
Perceived Importance by Council Members (ES2)	Strongly Disagree	26	6.1	6.1
	Disagree	63	14.7	20.8
	Neither Agree nor Disagree	109	25.5	46.3
	Agree	174	40.7	86.9
	Strongly Agree	56	13.1	100
	Total	428	100	
Council Meetings about PM (ES3)	Strongly Disagree	48	11.2	11.2
	Disagree	115	26.9	38.1
	Neither Agree nor Disagree	112	26.2	64.3
	Agree	123	28.7	93.0
	Strongly Agree	30	7.0	100
	Total	428	100	
Support of Citizens (ES4)	Strongly Disagree	57	13.3	13.3
	Disagree	120	28.0	41.4
	Neither Agree nor Disagree	107	25.0	66.4
	Agree	111	25.9	92.3
	Strongly Agree	33	7.7	100
	Total	428	100	
Citizen Interest in Performance Data (ES5)	Strongly Disagree	44	10.3	10.3
	Disagree	107	25.0	35.3
	Neither Agree nor Disagree	89	20.8	56.1
	Agree	137	32.0	88.1
	Strongly Agree	51	11.9	100
	Total	428	100	

First three questions regarding the external support are aimed at evaluating the support of council members. First question is a direct question asking directly the support from the council members in the municipality. More than half of the respondents (57%) reported their agreement on this statement. Although it is still high, it seems less than any of the organizational actors, even than the employees. Secondly, the respondents were asked whether the council members view performance measurement as an important aspect of decision making. Similar to the responses to the first question more than half of the respondents (53.7%) agreed or strongly agreed on this statement. Third question was aimed at understanding whether the interest of council members in performance measurement is reflected on the amount of council meetings in which performance measurement practices or performance data are discussed. Different from the first two questions, more respondents (38.1%) showed their disagreement on this statement than those who showed their agreement (35.7%).

Last two questions are about the support of citizens in the performance measurement practices of the municipality. Firstly, the respondents were asked directly about the support of citizens living in the municipal responsibility area. The respondents reported more disagreement (41.3%) than agreement (33.6%) on this statement. Lastly, the respondents were asked to show their opinions regarding the extent citizens show their interest to municipality's performance information. According to 43.9% of the responses citizens are interested in performance data, whereas 35.3% of the respondents oppose to this statement.

In general, it can be concluded that the responses for the indicators of external support do not indicate as clear a support as those of organizational support do. According to the responses, council members perceive performance measurement as useful and support the implementation, whereas performance measurement is not discussed much in the council

meetings. On the other hand, citizens' support for performance measurement is reported as being low. Yet, they still show their interest on performance information of the municipality.

5.1.3 Technical Capacity

Technical capacity is an endogenous mediating variable in the study model. It explains the extent the organization can implement the performance measurement systems. This capacity includes both the human resources capability and the technological capacity of the organization. In the survey, respondents were asked six questions about the level of technical capacity in their municipality. The frequency table for the indicators of technical capacity is given in the Table 7.

First two questions of technical capacity are about the adequacy of the number of staff and the information technology used in performance measurement steps. More respondents (46.9%) reported a problem regarding the adequacy of the number of staff than those (39.2%) reported that they have adequate staff. On the other hand, exactly half of the respondents state that they have adequate information technology, whereas 34.8% of the respondents respond negatively to this statement.

Later three questions are about the competencies of the staff for the development of high quality performance measures, timely collection of performance data, and the analysis of performance data. The positive and negative responses regarding these variables are more or less equal and around 40% range, meaning that around 40% of the municipalities reported problems in these areas, whereas the same amount of municipalities reported the adequacy in this regard.

Table 7 *Frequency and Percentage Distributions for Technical Capacity*

Indicator	Response	Frequency	Percentage	Cumulative Percentage
Staff (TC1)	Strongly Disagree	84	19.6	19.6
	Disagree	117	27.3	47.0
	Neither Agree nor Disagree	59	13.8	60.7
	Agree	126	29.4	90.2
	Strongly Agree	42	9.8	100
	Total	428	100	
Information Technology (TC2)	Strongly Disagree	51	11.9	11.9
	Disagree	98	22.9	34.8
	Neither Agree nor Disagree	65	15.2	50.0
	Agree	160	37.4	87.4
	Strongly Agree	54	12.6	100
	Total	428	100	
Competency in Performance Measure Development (TC3)	Strongly Disagree	75	17.5	17.5
	Disagree	111	25.9	43.5
	Neither Agree nor Disagree	72	16.8	60.3
	Agree	127	29.7	90.0
	Strongly Agree	43	10.0	100
	Total	428	100	
Competency in Timely Collection of Performance Data (TC4)	Strongly Disagree	70	16.4	16.4
	Disagree	105	24.5	40.9
	Neither Agree nor Disagree	62	14.5	55.4
	Agree	145	33.9	89.3
	Strongly Agree	46	10.7	100
	Total	428	100	
Competency in Performance Data Analysis (TC5)	Strongly Disagree	70	16.4	16.4
	Disagree	106	24.8	41.1
	Neither Agree nor Disagree	73	17.1	58.2
	Agree	132	30.8	89.0
	Strongly Agree	47	11.0	100
	Total	428	100	
Training (TC6)	Strongly Disagree	110	25.7	25.7
	Disagree	124	29.0	54.7
	Neither Agree nor Disagree	62	14.5	69.2
	Agree	94	22.0	91.1
	Strongly Agree	38	8.9	100
	Total	428	100	

The last question regarding technical capacity is whether the municipality has staff regularly attending to training activities, such as conferences, workshops, trainings, related to performance measurement. The responses show a clear negative answer to this question with 54.7% of the respondents disagreeing or strongly disagreeing as opposed to 30.9% agreeing or strongly agreeing.

In sum, it is difficult to draw a clear picture regarding the extent of technical capacity of Turkish municipalities based on the responses to this survey. Most of the respondents agree that their municipalities have adequate information technology and their staff is competent in timely collection of performance data, whereas most of the respondents report problems regarding the adequacy of staff number, competency of their staff in developing high quality performance measures, and the frequency their staff attend to the training activities. On the other hand, approximately same amount of respondents reported an agreement or disagreement of the competency of their staff regarding the analysis of performance data.

5.1.4 Quality of Performance Measures

Quality of performance measures is another endogenous variable in the study. As mentioned earlier, this study uses Bouckaert's (1993) model in assessing the quality of performance measures. According to this model, performance measures can be considered as effective, if they are valid, legitimate, and functional. The distributions of the responses regarding validity, legitimacy, and functionality of performance measures will be explained in this section.

5.1.4.1 Validity of Performance Measures. Validity of performance measures refers to the technical soundness of them. In the survey, respondents were asked six questions about

the level of validity of performance measures developed and used in their municipality. The frequency table for the indicators of validity is given in the Table 8.

Table 8 *Frequency and Percentage Distributions for Validity*

Indicator	Response	Frequency	Percentage	Cumulative Percentage
Derived from Missions/Goals (V1)	Strongly Disagree	42	9.8	9.8
	Disagree	55	12.9	22.7
	Neither Agree nor Disagree	69	16.1	38.8
	Agree	181	42.3	81.1
	Strongly Agree	81	18.9	100
	Total	428	100	
Derived from Service Standards (V2)	Strongly Disagree	40	9.3	9.3
	Disagree	59	13.8	23.1
	Neither Agree nor Disagree	69	16.1	39.3
	Agree	185	43.2	82.5
	Strongly Agree	75	17.5	100
	Total	428	100	
Focus on Importance (V3)	Strongly Disagree	31	7.2	7.2
	Disagree	50	11.7	18.9
	Neither Agree nor Disagree	79	18.5	37.4
	Agree	197	46.0	83.4
	Strongly Agree	71	16.6	100
	Total	428	100	
Being up to Date (V4)	Strongly Disagree	59	13.8	13.8
	Disagree	90	21.0	34.8
	Neither Agree nor Disagree	91	21.3	56.1
	Agree	132	30.8	86.9
	Strongly Agree	56	13.1	100
	Total	428	100	
Being Clear / Understandable (V5)	Strongly Disagree	41	9.6	9.6
	Disagree	55	12.9	22.4
	Neither Agree nor Disagree	92	21.5	43.9
	Agree	180	42.1	86.0
	Strongly Agree	60	14.0	100
	Total	428	100	
Measuring Performance over Time (V6)	Strongly Disagree	46	10.7	10.7
	Disagree	74	17.3	28.0
	Neither Agree nor Disagree	79	18.5	46.5
	Agree	169	39.5	86.0
	Strongly Agree	60	14.0	100
	Total	428	100	

First two questions regarding validity are whether performance measures are mostly developed from organizational missions/goals or from service standards. The respondents mostly supported both of these statements by an approximate 61% majority. A similar majority of the respondents also argued that their performance measures focus on what is important to measure rather than what data are available. On the other hand, the support for performance measures being up to date seems to be lower than the previous responses, yet there is more agreement (43.9%) than disagreement (34.8%) on this statement. Lastly, most of the respondents reported their agreement or strongly agreement on their performance measures being clear and measuring performance over time. In sum, the responses for the validity of performance measures reveal a support from the respondents about the validity of their performance measures.

5.1.4.2 Legitimacy of Performance Measures. Legitimacy of performance measures is about the positive perceptions of stakeholders regarding the performance measures. According to Bouckaert (1993), performance measurement is not only a technical issue, but also a motivational one. Involvement of employees and middle managers in the creation of performance measures can increase their commitment to performance measurement, effective implementation of which requires the approval of these groups (Bouckaert, 1993). In the survey, respondents were asked six questions about the legitimacy of performance measures developed and used in their municipality. The frequency table for the indicators of legitimacy is given in the Table 9.

Table 9 *Frequency and Percentage Distributions for Legitimacy*

Indicator	Response	Frequency	Percentage	Cumulative Percentage
Involvement of Managers (L1)	Strongly Disagree	44	10.3	10.3
	Disagree	77	18	28.3
	Neither Agree nor Disagree	77	18	46.3
	Agree	180	42.1	88.3
	Strongly Agree	50	11.7	100
	Total	428	100	
Involvement of Employees (L2)	Strongly Disagree	50	11.7	11.7
	Disagree	110	25.7	37.4
	Neither Agree nor Disagree	101	23.6	61
	Agree	139	32.5	93.5
	Strongly Agree	28	6.5	100
	Total	428	100	
Informing Council Members (L3)	Strongly Disagree	54	12.6	12.6
	Disagree	94	22.0	34.6
	Neither Agree nor Disagree	92	21.5	56.1
	Agree	156	36.4	92.5
	Strongly Agree	32	7.5	100
	Total	428	100	
Perceived Usefulness by Elected Officials (L4)	Strongly Disagree	41	9.6	9.6
	Disagree	52	12.1	21.7
	Neither Agree nor Disagree	107	25.0	46.7
	Agree	176	41.1	87.9
	Strongly Agree	52	12.1	100
	Total	428	100	
Perceived Usefulness by Managers (L5)	Strongly Disagree	35	8.2	8.2
	Disagree	42	9.8	18.0
	Neither Agree nor Disagree	89	20.8	38.8
	Agree	204	47.7	86.4
	Strongly Agree	58	13.6	100
	Total	428	100	
Perceived Usefulness by Employees (L6)	Strongly Disagree	41	9.6	9.6
	Disagree	57	13.3	22.9
	Neither Agree nor Disagree	124	29.0	51.9
	Agree	166	38.8	90.7
	Strongly Agree	40	9.3	100
	Total	428	100	

For the legitimacy there are two groups of questions. First three questions concern the involvement of several actors in the development of performance measures. More than half of the respondents (53.8%) agree or strongly agree that most of the managers involve in the development process. On the other hand, there is only a slight difference (39% to 37.4%) between those who agrees that employees involve in the process and those who disagrees to that. The responses for keeping the council members informed about the process seem to take a middle ground between the first two questions with 43.9% agreeing and 34.6% disagreeing.

Second group of questions are concerned with the perceived usefulness of the created performance measures. In that sense, perceptions of managers, employees, and council members are asked in the survey. Most of the respondents reported a positive perception of these groups regarding the usefulness of performance measures rather than a negative one. Yet, the perception of usefulness for employees seems to be lower (48.1%) than that for managers (61.3%).

5.1.4.3 Functionality of Performance Measures. The last dimension of the quality of a performance measure is its functionality. Functionality, in this sense, refers to the benefit creating potential of the measures. In the survey, respondents were asked five questions about the functionality of performance measures developed and used in their municipality. The frequency table for the indicators of functionality is given in the Table 10.

In the survey, the respondents were asked whether the performance measures developed in their municipality have the potential to improve service quality, decision-making capacity, employee motivation, organizational learning, and communication of managers with elected officials. An overwhelming majority of the respondents, ranging from

65.2% to 75.7%, agreed or strongly agreed to these five positive statements related to the functionality of performance measures developed in their municipality.

Table 10 *Frequency and Percentage Distributions for Functionality*

Indicator	Response	Frequency	Percentage	Cumulative Percentage
Potential for Service Quality Improvement (F1)	Strongly Disagree	26	6.1	6.1
	Disagree	34	7.9	14.0
	Neither Agree nor Disagree	68	15.9	29.9
	Agree	207	48.4	78.3
	Strongly Agree	93	21.7	100
	Total	428	100	
Potential for Decision-making Capacity Improvement (F2)	Strongly Disagree	25	5.8	5.8
	Disagree	29	6.8	12.6
	Neither Agree nor Disagree	50	11.7	24.3
	Agree	228	53.3	77.6
	Strongly Agree	96	22.4	100
	Total	428	100	
Potential for Increasing Employee Motivation (F3)	Strongly Disagree	23	5.4	5.4
	Disagree	31	7.2	12.6
	Neither Agree nor Disagree	75	17.5	30.1
	Agree	209	48.8	79.0
	Strongly Agree	90	21.0	100
	Total	428	100	
Potential for Stimulating Organizational Learning (F4)	Strongly Disagree	25	5.8	5.8
	Disagree	30	7.0	12.9
	Neither Agree nor Disagree	67	15.7	28.5
	Agree	215	50.2	78.7
	Strongly Agree	91	21.3	100
	Total	428	100	
Potential for Improving External Communication (F5)	Strongly Disagree	27	6.3	6.3
	Disagree	33	7.7	14.0
	Neither Agree nor Disagree	89	20.8	34.8
	Agree	205	47.9	82.7
	Strongly Agree	74	17.3	100
	Total	428	100	

5.1.5 Effectiveness of Performance Measurement

Effectiveness of performance measurement is the endogenous variable of the study. It is related to the implementation phase of performance measurement and concerns mainly the effects of performance measurement. In the survey, respondents were asked ten questions about the effectiveness of performance measurement used in their municipality. The frequency table for the indicators of effectiveness of performance measurement is given in the Table 11.

Table 11 *Frequency and Percentage Distributions for Effectiveness of Performance Measurement*

Indicator	Response	Frequency	Percentage	Cumulative Percentage
Improvement in Productivity (E1)	Strongly Disagree	24	5,6	5,6
	Disagree	34	7,9	13,6
	Neither Agree nor Disagree	70	16,4	29,9
	Agree	220	51,4	81,3
	Strongly Agree	80	18,7	100
	Total	428	100	
Improvement in Service Quality (E2)	Strongly Disagree	23	5,4	5,4
	Disagree	28	6,5	11,9
	Neither Agree nor Disagree	64	15,0	26,9
	Agree	217	50,7	77,6
	Strongly Agree	96	22,4	100
	Total	428	100	
Increase in Employee Motivation (E3)	Strongly Disagree	23	5,4	5,4
	Disagree	33	7,7	13,1
	Neither Agree nor Disagree	83	19,4	32,5
	Agree	209	48,8	81,3
	Strongly Agree	80	18,7	100
	Total	428	100	
Stimulation of Organizational Learning (E4)	Strongly Disagree	25	5,8	5,8
	Disagree	24	5,6	11,4
	Neither Agree nor Disagree	79	18,5	29,9
	Agree	219	51,2	81,1
	Strongly Agree	81	18,9	100
	Total	428	100	
Improved Relationship	Strongly Disagree	24	5,6	5,6

Indicator	Response	Frequency	Percentage	Cumulative Percentage
with Community (E5)	Disagree	31	7,2	12,9
	Neither Agree nor Disagree	88	20,6	33,4
	Agree	208	48,6	82,0
	Strongly Agree	77	18,0	100
	Total	428	100	
Cost Reduction (E6)	Strongly Disagree	26	6,1	6,1
	Disagree	31	7,2	13,3
	Neither Agree nor Disagree	96	22,4	35,7
	Agree	191	44,6	80,4
	Strongly Agree	84	19,6	100
	Total	428	100	
Better Identification of Problems (E7)	Strongly Disagree	22	5,1	5,1
	Disagree	25	5,8	11,0
	Neither Agree nor Disagree	59	13,8	24,8
	Agree	227	53,0	77,8
	Strongly Agree	95	22,2	100
	Total	428	100	
Better Solution of Problems (E8)	Strongly Disagree	20	4,7	4,7
	Disagree	21	4,9	9,6
	Neither Agree nor Disagree	61	14,3	23,8
	Agree	236	55,1	79,0
	Strongly Agree	90	21,0	100
	Total	428	100	
Better Decision-making (E9)	Strongly Disagree	22	5,1	5,1
	Disagree	19	4,4	9,6
	Neither Agree nor Disagree	63	14,7	24,3
	Agree	224	52,3	76,6
	Strongly Agree	100	23,4	100
	Total	428	100	
Better Communication with Elected Officials (E10)	Strongly Disagree	22	5,1	5,1
	Disagree	26	6,1	11,2
	Neither Agree nor Disagree	78	18,2	29,4
	Agree	217	50,7	80,1
	Strongly Agree	85	19,9	100
	Total	428	100	

In the survey, the respondents were firstly asked whether the use of performance measurement in their municipality improved productivity and service quality, increased

employee motivation, stimulated organizational learning, improved relations with the community, helped managers to identify the problems and create solutions for these problems, facilitated better decisions, and contributed to the managerial communication with elected officials. Among these indicators, the statement with the least positive response rate (64.2%) was the cost reduction benefit of performance measurement. On the other hand, developing better solutions to managerial and operational problems received the highest number of positive statements (76.1%) among all indicators of effectiveness of performance measurement. Most of the respondents agreed or strongly agreed to these ten questions indicating a clear support for the effectiveness of performance measurement in their municipalities.

5.2 Correlation Analysis

Correlation analysis is useful to examine the correlations among the indicators of the latent variables and to detect if multicollinearity exists among them. Multicollinearity occurs if two indicators of a latent variable are highly correlated to each other, which means that they mainly measure the same thing (Kline, 2011). As a result, using both of these indicators is redundant and one of them should be removed from the model.

In the study, Spearman's rank order correlation, which is mostly known as Spearman's rho, is used to examine the correlations among the observed variables, since the observed variables in the study are ordinal variables and their correlation is better examined by this method (Kline, 2011). The scholars mention several thresholds, ranging from .7 to .9, for deciding the presence of multicollinearity among the variables. In this study, .85 was set as the threshold for detecting multicollinearity.

The indicators of each latent construct are assessed together. The correlation matrixes of the variables are given in Appendix D. The examination of correlation among the indicators of organizational support reveals that the correlation coefficient values for the indicator pairs range from .409 to .768. The indicator pair with the lowest correlation is *support of mayor* (OS1) and *support of employees* (OS4). On the other hand, correlation between *support of mayor* (OS1) and *support of top-managers* (OS2) is at the highest level among the indicators. All of the indicators have statistically significant correlation at .01 level, however none of them exceeds the predetermined threshold of .85. Therefore, none of the indicators will be excluded from the model.

According to the correlation matrix for the indicators of external support, all of the indicators show statistically significant correlation at .01 level and correlation coefficients range from .454 to .791. The indicators with the lowest inter-correlation are *support of council members* (ES1) and *citizen interest in performance data* (ES5). Not surprisingly, s*upport of council members* (ES1) and *perceived importance by council members* (ES2) show the highest inter-correlation. However, even this coefficient does not exceed the predetermined threshold for detecting multicollinearity. As a result, none of the indicators were needed to be excluded from the model.

The third latent variable for the correlation analysis is technical capacity. Compared to the first two variables, the indicators of this variable show higher correlations among each other. Correlation coefficients range from .573 to .890. Yet, there is only one correlation which exceeds the threshold. It is the correlation among *competency in timely collection of performance data* (TC4) and *competency in performance data analysis* (TC5). This means that respondents who think that their municipality has competent staff for collecting performance data in a timely manner are more likely to think that they have competent staff

for analyzing this performance data or vice versa. Since the correlation exceeds the .85 threshold, one of the indicators should be excluded from the model. Based on the results of a preliminary confirmatory factor analysis (see Figure 30 in Appendix E), TC5 was excluded from the model, since it had a slightly less importance for the latent construct of technical capacity than TC4 had.

Following three variables, correlations among whose indicators are examined, are the dimensions of high quality performance measures, namely validity, legitimacy, and functionality. Firstly, correlation among the indicators of validity was examined. All of the correlations among the indicators of validity show statistical significance at .01 level. Correlation coefficients of the indicators range from .701 to .906. The only correlation which exceeds the threshold set for detecting multicollinearity is the correlation between *derived from missions/goals* (V1) and *derived from service standards* (V2). As a result, one of the indicators should be excluded from the model. Based on the results of a preliminary confirmatory factor analysis (see Figure 31 in Appendix E), V2 was excluded from the model, since it had a slightly less importance for the latent construct of validity than V1.

Although the correlations among indicators of legitimacy are lower than those of the validity, still all of the correlations are statistically significant at .01 level. Correlation coefficients range from .495 to .802. The indicators with the lowest inter-correlation are *involvement of employees* (L2) and *perceived usefulness by elected officials* (L4). On the other hand, *perceived usefulness by elected officials* (L4) and *perceived usefulness by managers* (L5) show the highest inter-correlation. Since there is not any correlation over the threshold of .85, none of the indicators were excluded from the model.

An examination of the correlation matrix of functionality reveals that most of the correlations are situated around .8 and they are ranging from .783 and .840. The correlation between *potential for stimulating organizational learning* (F4) and *potential for improving external communication* (F5) has the highest value among all indicators. However, even this value is lower than the threshold, meaning that none of the indicators were needed to be excluded from the model.

The last construct for which correlation analysis is run is effectiveness of performance measurement, which is also the endogenous variable of the study. All of the correlations among the indicators are statistically significant at .01 level and the coefficients range from .671 to .921. A further examination of the Table 38 (in Appendix D) reveals that 3 pairs of indicators have correlations higher than the threshold level for multicollinearity. These pairs are: *improvement in productivity* (E1) and *improvement in service quality* (E2) with .921, *improvement in service quality* (E2) and stimulation *of organizational learning* (E4) with .851, and *better identification of problems* (E7) and *better solution of problems* (E8) with .885. Moreover, correlation between *better solution of problems* (E8) and *better decision-making* (E9) is very close to the threshold with .848. In order to eliminate multicollinearity problem in the first two pairs, *improvement in service quality* (E2) was excluded from the model. Similarly, *better solution of problems* (E8) was excluded from the model to eliminate the problem for the last two pairs.

After examining the descriptive characteristics of the data and the correlations among the indicators of latent variables, the next is step is to test the proposed measurement models by using confirmatory factor analysis, and to make necessary revisions to validate the measurement models, which will be used as the basis of covariance structure model of the study.

5.3 Confirmatory Factor Analysis

Following the descriptive analysis, confirmatory factor analysis (CFA) was conducted by using AMOS 22 software to evaluate and validate the measurement models of the latent study variables. Since the latent constructs cannot be measured directly, measurement models composed of several indicators are utilized to measure those constructs (Bryne, 2010). However, it is important to check the validity of this measurement model before proceeding to the analysis of covariance structure model. This is done by using confirmatory factor analysis (Wan, 2002; Bryne, 2010).

The validation of the proposed model took place in two stages. In the first stage, each measurement model was evaluated and, if needed, revised by using a three-step method proposed by Wan (2002). These steps are; 1) checking the appropriateness of the indicators, 2) checking the overall model fit, 3) revising the model. These steps have been explained in more depth in the previous chapter. In the second stage, the validated measurement models was integrated into the covariance structure model (CSM) and the CSM was evaluated and revised by using the goodness of fit statistics and modification indices.

In the study, there are five main latent variables; two exogenous (organizational support and external support), two endogenous mediating (technical capacity and quality of performance measures), and one endogenous (effectiveness of performance measurement) variable. However, quality of performance measures is a second-order variable with three first-order latent variables (validity, legitimacy, and functionality), whose measurement models are also needed to be validated. As a result, a total of eight measurement models will be evaluated and validated in this section.

5.3.1 Organizational Support

The proposed measurement model for organizational support has seven indicators (see Table 1 and Figure 5). Since none of the indicators had high correlation among each other, all of the indicators in the proposed model were kept in the generic model, which is shown in Figure 11.

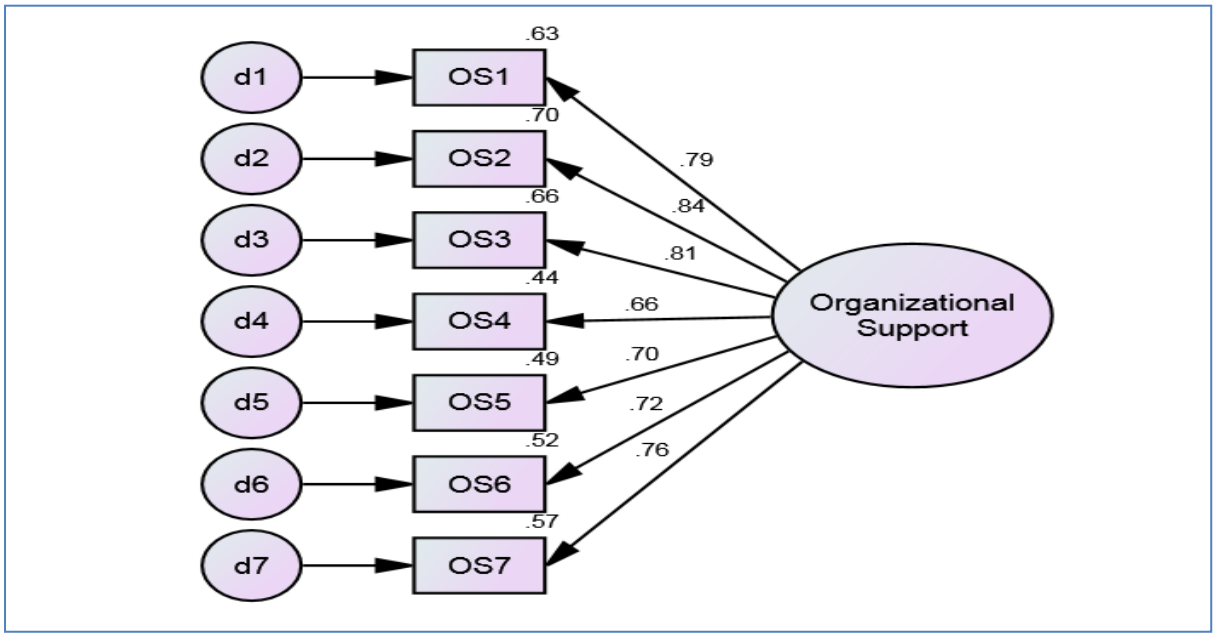

Figure 11. Generic Measurement Model for Organizational Support

First step of the analysis is to examine the significance and strength of factor loadings, which is the association between the indicators and their latent constructs. Examination of Table 12 shows that all of the indicators of organizational support have critical ratios higher than 1.96 and p values lower than .05, which means that factor loadings for all indicators show statistical significance. Secondly, the strengths of the factor loadings are examined by looking at their standardized regression weight values. Since none of the factor loadings had values lower than .30, which is the threshold level of the study for an indicator to be considered as important, all of the indicators were kept in the model.

However, examination of the goodness of fit statistics (Table 13), none of which showed a good fit, revealed that the proposed study model did not fit the data. As a result, a necessity for the modification of the model occurred. Modification of the model can be done both by dropping the indicators with statistically insignificant or unimportant factor loadings, and correlating the measurement errors. Since none of the factor loadings are neither statistical insignificant nor unimportant, modification of the model was done by correlating the measurement errors, for which examination of the modification indices, which are generated by the AMOS software, is important for figuring out which correlated errors should be freely estimated in order to reduce the chi-square value and make the model fit better to the data at hand. Beginning from the highest modification indices, correlations among several measurement errors were identified and revised measurement models for the study variables were built. The revised measurement model for organizational support is presented in Figure 12.

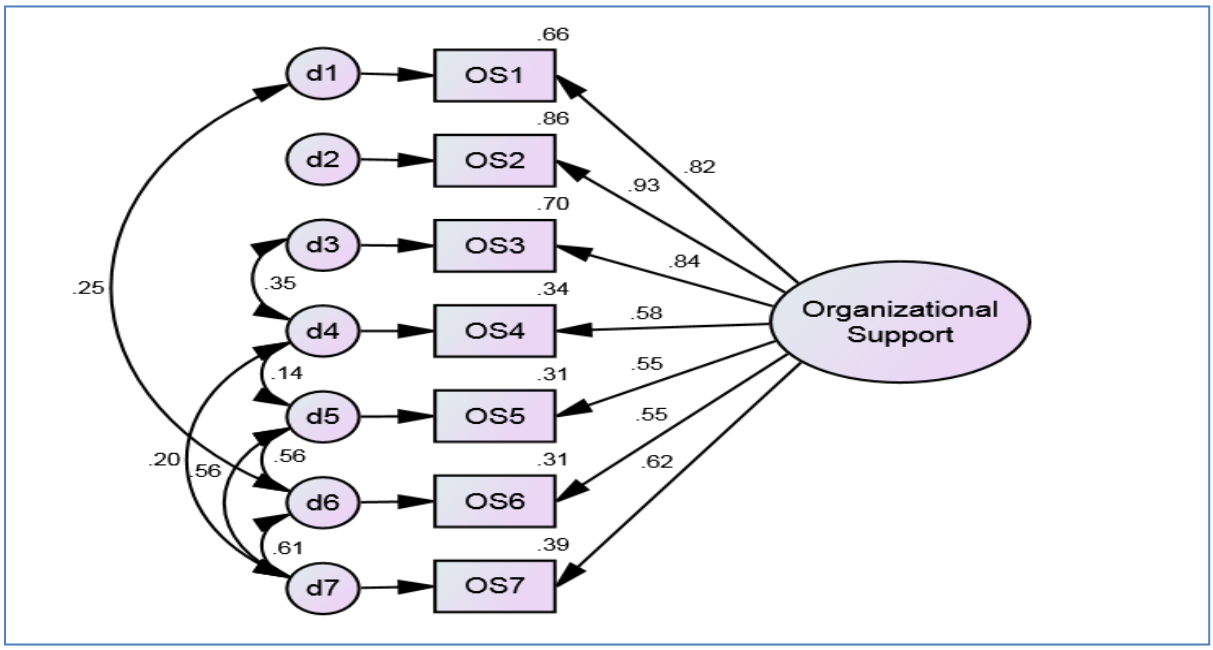

Figure 12. Revised Measurement Model for Organizational Support

Parameter estimates of both the generic model and the revised model are given in Table 12. The examination of the table shows that all of the indicators and the correlations among the measurement errors have statistical significance. Moreover, all of the factor loadings exceeded the predetermined criteria of .30. As a result, all of the indicators were decided to be kept in the model.

Table 12 *Parameter Estimates for Organizational Support*

	Generic Model					Revised Model				
Indicator	URW	SRW	SE	CR	P	URW	SRW	SE	CR	P
OS1	.831	.794	.055	15.154	***	1.082	.815	.092	11.804	***
OS2	.995	.839	.062	15.930	***	1.393	.925	.112	12.420	***
OS3	.919	.810	.059	15.439	***	1.207	.839	.101	11.980	***
OS4	.832	.663	.065	12.800	***	.924	.581	.090	10.268	***
OS5	1.000	.697				1.000	.554			
OS6	1.028	.721	.074	13.851	***	.999	.555	.072	13.944	***
OS7	1.022	.756	.071	14.482	***	1.060	.625	.071	14.992	***
d1<-->d6						.115	.248	.019	5.983	***
d4<--> d3						.142	.349	.024	5.992	***
d4<--> d7						.140	.203	.026	5.356	***
d5<--> d6						.508	.562	.050	10.147	***
d5<--> d7						.444	.556	.046	9.617	***
d5<--> d4						.109	.139	.030	3.615	***
d6<--> d7						.487	.611	.045	10.795	***

Note: URW = Unstandardized Regression Weight; SRW = Standardized Regression Weight; SE = Standard Error; CR = Critical Ratio; *** = Correlation is significant at .01 level

Goodness of fit statistics of both generic and revised models are given in Table 13. The table indicates that all of the statistics show poor fit for the generic model, but they all show excellent fit for the revised model. All of the indices met even all of the possible criteria for an excellent model fit mentioned in the literature.

Table 13 *Goodness of Fit Statistics for Organizational Support*

Fit Index	Criteria	Generic Model	Revised Model
Chi-Square (CMIN)	Smaller the better	460.788	13.953
Chi-Square related p value	≥ .05	.000	.052
Chi-square / Degree of Freedom (CMIN /df)	≤ 4	32.913	1.993
Tucker Lewis Index (TLI)	≥ .90	.668	.990
Comparative Fit Index (CFI)	≥ .90	.778	.997
Root Mean Square Error of Approximation (RMSEA)	≤ .08	.273	.048
Hoelter's Critical N	≥ 75	22	431

After achieving the model fit, the next step is to evaluate and interpret the importance of indicators in the measurement model. This process is conducted by looking at their standardized regression weight values. These weights make it possible to compare effects of different indicators on the latent variable. According to Table 12, *support of top-managers* (OS2) is the most important indicator with a standardized regression weight of .925, whereas *special meetings* (OS5) and *mayor's communication about performance measurement* (OS6) are the least important indicators with values of .554 and .555 respectively.

Based on the examination of the tables, it can be concluded that measurement model for organizational support is validated and ready for further analysis of internal consistency and covariance structure model.

5.3.2 External Support

The proposed measurement model for external support had five indicators (see Table 1 and Figure 6). Since none of the indicators had high correlation among each other, all of the indicators in the proposed model were kept in the generic model, which is shown in Figure 13.

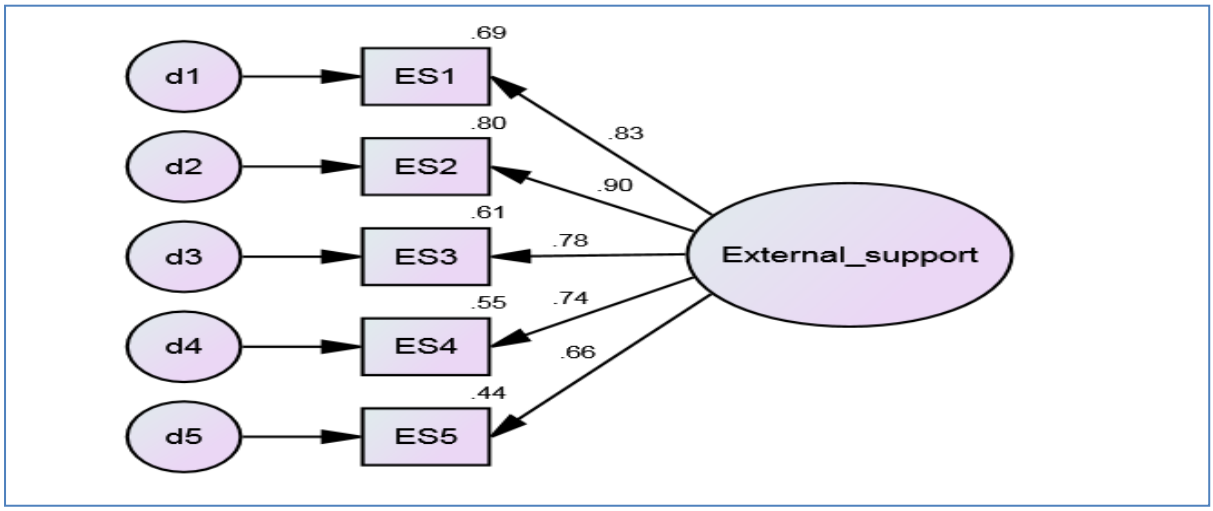

Figure 13. Generic Measurement Model for External Support

After conducting the confirmatory factor analysis, firstly, the significance and strength of the factor loadings, which are presented in Table 14, were examined. All of the indicators of external support have critical ratios higher than 1.96, p values lower than .05 and standardized regression weights higher than .30. As a result, none of the indicators were needed to be excluded from the model. However, according to the goodness of fit statistics shown in Table 15, the model did not fit well to the data. Firstly, correlating the measurement errors was tried, but getting an acceptable model fit could not be achieved. Therefore, the measurement model was revised by excluding the least important indicator in the model, which is *citizen interest in performance data* with a standardized regression weight of .664.

Confirmatory factor analysis was run again with the new model. The first revised model is presented in Figure 14.

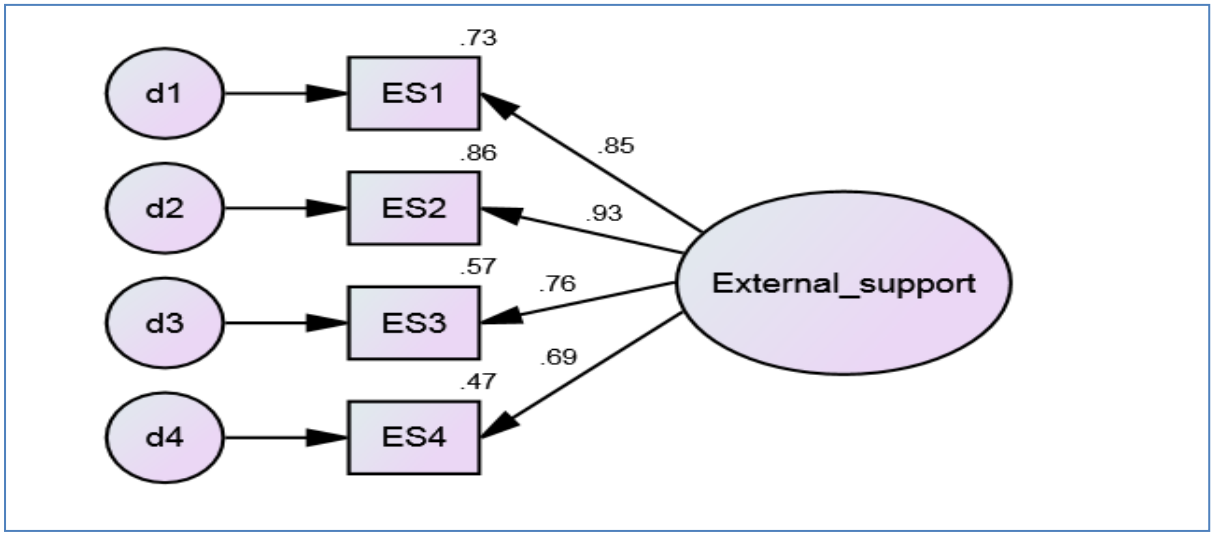

Figure 14. First Revised Measurement Model for External Support

Examination of parameter estimates, presented in Table 14, shows that factor loadings of all indicators are statistically significant and their standardized regression weights are higher than .30. Therefore, none of these indicators were removed from the model. However, all of the goodness of fit indices (Table 15), but CFI, show a poor model fit, meaning that that the model should be revised. For the revision, measurement errors were correlated one at a time according to the modification indices generated by AMOS. Only one correlation, which is between E3 and E4, sufficed for an excellent model fit. Final revised measurement model of external support is presented in Figure 15.

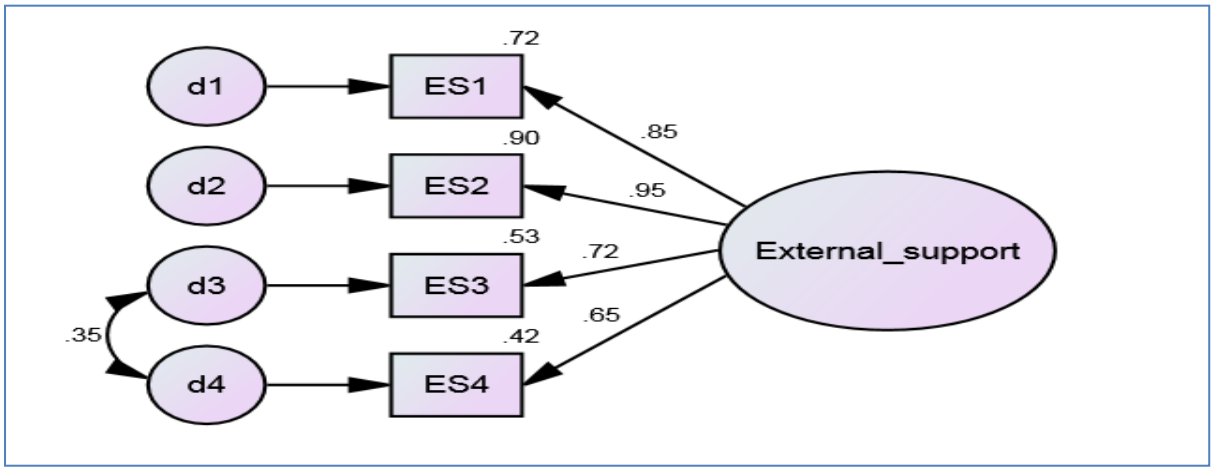

Figure 15. Final Revised Measurement Model for External Support

Parameter estimates of the generic model and both of the revised models are given in Table 14. The examination of the table shows that all of the indicators and the correlations among the measurement errors have statistical significance. Moreover, all of the factor loadings exceeded the predetermined criteria of .30. As a result, all of the indicators were decided to be kept in the revised model.

Table 15 shows the goodness of fit statistics of the generic and both of the revised models. The table indicates that all of the statistics for the generic model do not satisfy the predetermined criteria for the model fit. Similar problem occurs for the first revised model, in which only CFI reaches to its threshold level. On the other hand, the statistics for the final revised model show excellent fit.

After achieving the model fit, the importance of indicators in the revised measurement model can be evaluated. According to Table 14, all of the indicators have standardized regression weights ranging from .652 to .948, meaning that they all are important factors influencing the external support of the municipality. According to the table, p*erceived importance by council members* (ES2) is the most important indicator, whereas s*upport of citizens* (ES4) is the least important one.

Table 14 *Parameter Estimates for External Support*

	Generic Model					First Revised Model					Final Revised Model				
Indicator	URW	SRW	SE	CR	P	URW	SRW	SE	CR	P	URW	SRW	SE	CR	P
ES1	1.000	.831				1.000	.852				1.000	.848			
ES2	1.106	.895	.050	21.995	***	1.116	.926	.048	23.385	***	1.148	.948	.052	22.192	***
ES3	1.017	.783	.055	18.451	***	.959	.757	.053	18.247	***	.922	.725	.053	17.248	***
ES4	.995	.743	.058	17.147	***	.899	.688	.056	15.921	***	.855	.652	.057	14.889	***
ES5	.917	.664	.062	14.812	***	-	-	-	-	-	-	-	-	-	-
d3<--> d4											.241	.349	.040	5.961	***

Note: URW = Unstandardized Regression Weight; SRW = Standardized Regression Weight; SE = Standard Error; CR = Critical Ratio;
*** = Correlation is significant at .01 level

Table 15 *Goodness of Fit Statistics for External Support*

Fit Index	Criteria	Generic Model	First Revised Model	Final Revised Model
Chi-Square (CMIN)	Smaller the better	155.262	46.923	1.242
Chi-Square related p value	$\geq .05$	.000	.000	.265
Chi-square / Degree of Freedom (CMIN /df)	≤ 4	31.052	23.462	1.242
Tucker Lewis Index (TLI)	$\geq .90$	.771	.866	.999
Comparative Fit Index (CFI)	$\geq .90$	.886	.955	1.000
Root Mean Square Error of Approximation (RMSEA)	$\leq .08$	.265	.229	.024
Hoelter's Critical N	≥ 75	31	55	1321

5.3.3 Technical Capacity

The proposed measurement model for technical capacity had six indicators (see Table 1 and Figure 7). However, correlation analysis revealed that there is high correlation (.890) between two indicators, namely *competency in timely collection of performance data* (TC4) and *competency in performance data analysis* (TC5), which signs the high possibility of multicollinearity, and therefore, TC5 was excluded from the model. As a result, generic model for technical capacity consists of five indicators. The model is presented in Figure 16.

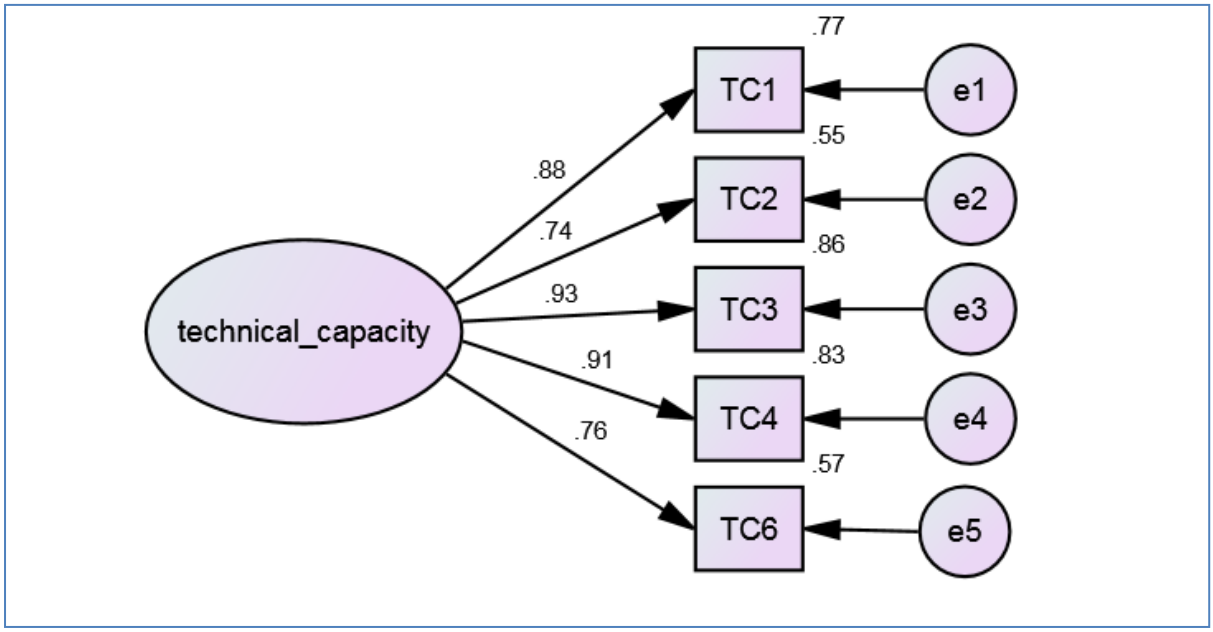

Figure 16. Generic Measurement Model for Technical Capacity

Firstly, the significance and strength of factor loadings were examined. Examination of Table 16 shows that all of the indicators of organizational support have critical ratios higher than 1.96 and p values lower than .05, which means that factor loadings for all indicators show statistical significance. Moreover, all of the factor loadings are found to be important, since they exceed the predetermined criteria of .30. Therefore, all of the indicators were decided to be kept in the model.

On the other hand, examination of the goodness of fit statistics in Table 17 revealed that while some of the indices, like CFI, TLI, and Hoelter Index, showed a good fit, the others were below the predetermined criteria. Therefore, a revision was considered to be helpful in order to achieve a better fit for the measurement model. For the revision, measurement errors of the indicators were correlated one at a time according to the modification indices generated by AMOS. Measurement errors of two pairs of indicators, between E1 - E2 and between E4-E6, were correlated to reach a better model fit. The revised measurement model for technical capacity is presented in Figure 17.

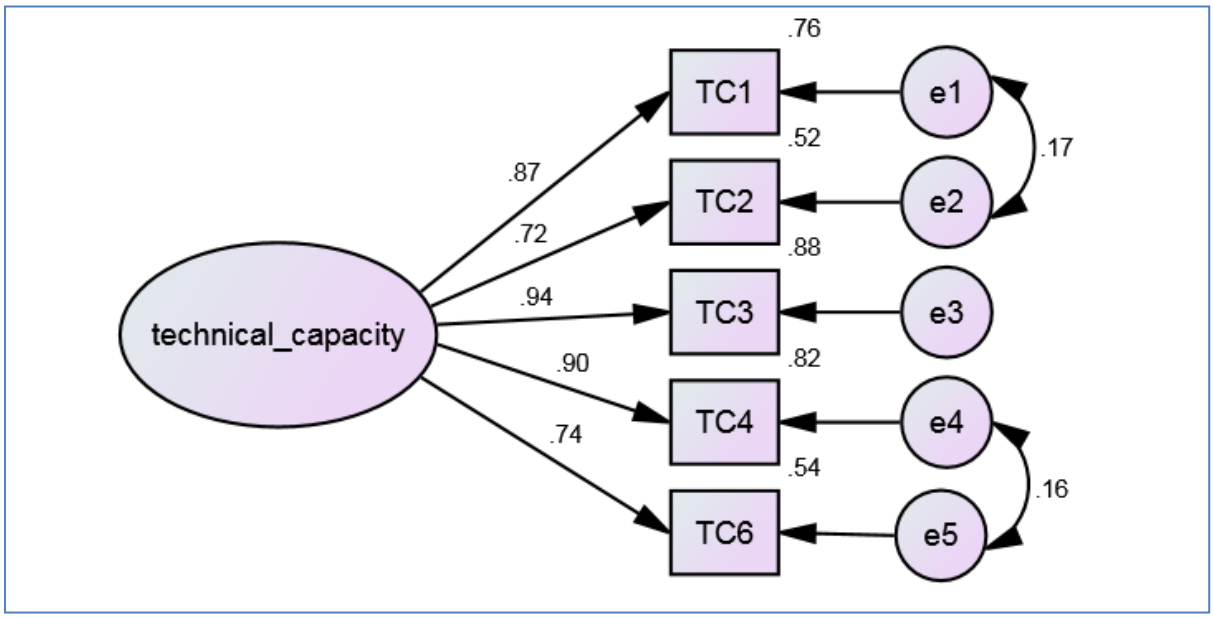

Figure 17. Revised Measurement Model for Technical Capacity

Parameter estimates of both the generic model and the revised model are given in Table 16. The examination of the table shows that all of the indicators and the correlations among the measurement errors have statistical significance. Moreover, all of the factor loadings exceeded .30. As a result, all of the indicators are decided to be kept in the revised model.

Table 16 *Parameter Estimates for Technical Capacity*

	Generic Model					Revised Model				
Indicator	URW	SRW	SE	CR	P	URW	SRW	SE	CR	P
TC1	1.000	.879				1.000	.872			
TC2	.802	.739	.043	18.749	***	.792	.724	.041	19.473	***
TC3	1.033	.926	.036	28.561	***	1.052	.936	.037	28.100	***
TC4	1.020	.909	.037	27.485	***	1.022	.904	.039	26.426	***
TC6	.864	.756	.044	19.450	***	.849	.738	.046	18.264	***
e1<--> e2						.093	.169	.033	2.810	.005
e4<--> e5						.080	.164	.032	2.487	.013

Note: URW = Unstandardized Regression Weight; SRW = Standardized Regression Weight; SE = Standard Error; CR = Critical Ratio; *** = Correlation is significant at .01 level

Goodness of fit statistics of both generic and revised models are given in Table 17. The table indicates that some of the statistics show poor fit for the generic model, but they all show excellent fit for the revised model. All of the indices met even all of the possible criteria for an excellent model fit mentioned in the literature.

Table 17 *Goodness of Fit Statistics for Technical Capacity*

Fit Index	Criteria	Generic Model	Revised Model
Chi-Square (CMIN)	Smaller the better	21.036	5.190
Chi-Square related p value	≥ .05	.001	.158
Chi-square / Degree of Freedom (CMIN /df)	≤ 4	4.207	1.730
Tucker Lewis Index (TLI)	≥ .90	.981	.996
Comparative Fit Index (CFI)	≥ .90	.991	.999
Root Mean Square Error of Approximation (RMSEA)	≤ .08	.087	.041
Hoelter's Critical N	≥ 75	225	643

After achieving the model fit, the next step is to evaluate and interpret the significance and importance of indicators in the measurement model. The examination of Table 14 shows that all of the indicators have standardized regression weights ranging from .724 to .936, meaning that they all are important factors influencing the technical capacity of the municipality in performance measurement. According to the table, *competency in performance measure development* (TC3) is the most important indicator with a standardized regression weight of .936, whereas *information technology* (TC2) is the least important indicator of technical capacity with a value of .724.

5.3.4 Quality of Performance Measures

Quality of performance measures is an endogenous mediating second-order variable, which has three dimensions; validity, legitimacy, and functionality. Firstly, measurement models of these three first-order variables will be validated. Then, these models will be combined together to form the measurement model of the quality of performance measures.

5.3.4.1 Validity. The proposed measurement model for validity had six indicators (see Table 1 and Figure 8). However, correlation analysis revealed that there is high correlation (.906) between two indicators, namely between *derived from missions/goals* (V1) and *derived from service standards* (V2), which flags a high risk for multicollinearity. Therefore, V2 was excluded from the model. As a result, generic model for validity consisted of five indicators. The model is presented in Figure 18.

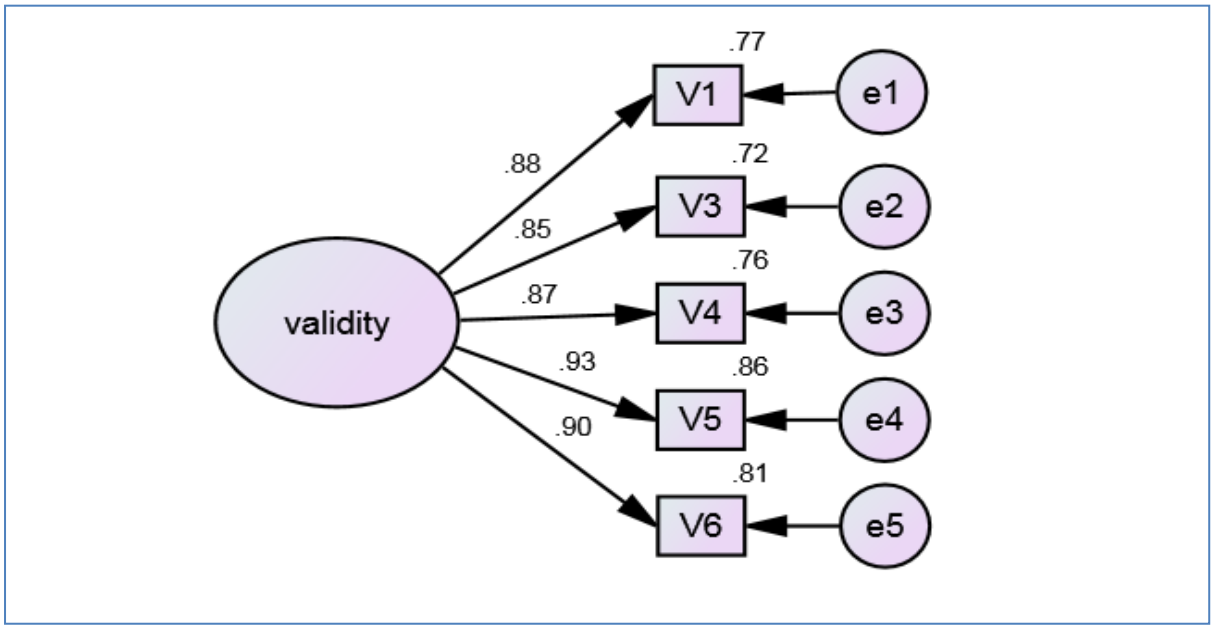

Figure 18. Generic Measurement Model for Validity

Firstly, the significance and strength of factor loadings were examined. Examination of Table 18 shows that all of the indicators of validity have critical ratios higher than 1.96 and p values lower than .05, which means that factor loadings for all indicators show statistical significance. Moreover, all of the factor loadings are found to be important, since they exceed the predetermined criteria of .30. Therefore, all of the indicators were decided to be kept in the model.

On the other hand, examination of the goodness of fit statistics in Table 19 revealed that some of the indices, like CFI, TLI, and Hoelter Index, showed a good fit, whereas the others did not satisfy the threshold levels determined for this study. Therefore, a revision was done in the model to achieve a better fit. For the revision, measurement errors of two indicators, E1 and E3, were correlated. The revised measurement model for validity is presented in Figure 19.

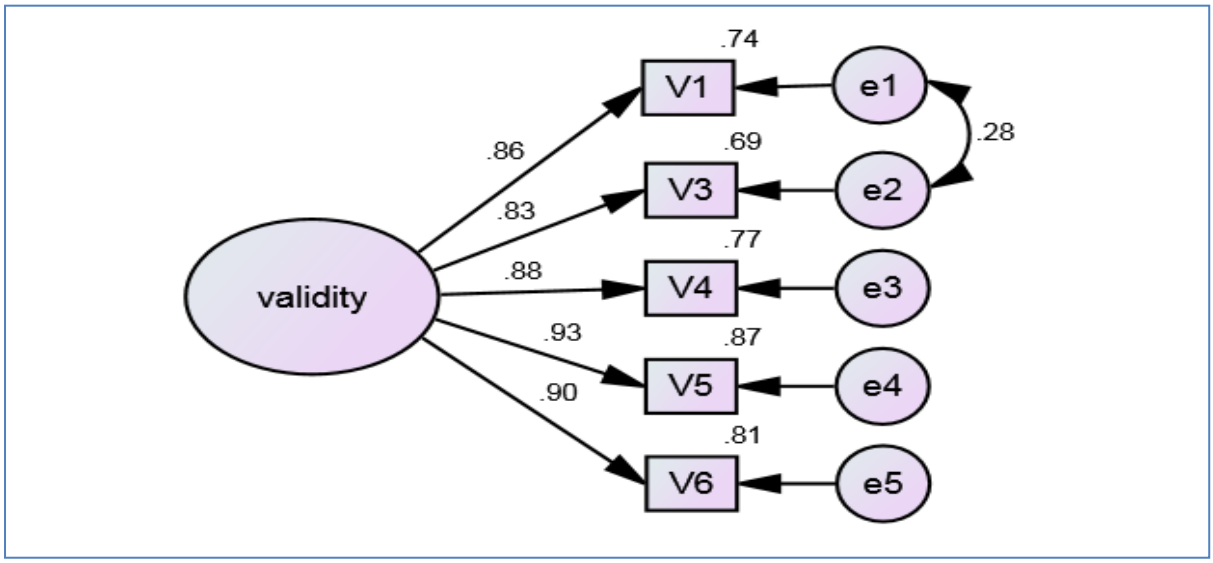

Figure 19. Revised Measurement Model for Validity

Parameter estimates of both the generic model and the revised model are given in Table 18. The examination of the table shows that all of the indicators and the correlations among the measurement errors have statistical significance. Moreover, all of the factor loadings exceeded .30, showing that they are important factors for the validity of performance measures. As a result, all of the indicators are decided to be kept in the revised model.

Table 18 *Parameter Estimates for Validity*

	Generic Model					Revised Model				
Indicator	URW	SRW	SE	CR	P	URW	SRW	SE	CR	P
V1	1.000	.877				1.000	.861			
V3	.893	.851	.037	24.170	***	.891	.833	.034	26.520	***
V4	1.033	.873	.041	25.441	***	1.057	.877	.043	24.589	***
V5	1.012	.928	.035	29.056	***	1.038	.934	.037	27.794	***
V6	1.025	.898	.038	27.015	***	1.048	.901	.040	25.909	***
e1<--> e2						.109	.285	.024	4.597	***

Note: URW = Unstandardized Regression Weight; SRW = Standardized Regression Weight; SE = Standard Error; CR = Critical Ratio; *** = Correlation is significant at .01 level

Goodness of fit statistics of both generic and revised models are given in Table 19. The table indicates that some of the statistics show poor fit for the generic model, but they all show excellent fit for the revised model.

Table 19 *Goodness of Fit Statistics for Validity*

Fit Index	Criteria	Generic Model	Revised Model
Chi-Square (CMIN)	Smaller the better	32.754	6.791
Chi-Square related p value	≥ .05	.000	.147
Chi-square / Degree of Freedom (CMIN /df)	≤ 4	6.551	1.698
Tucker Lewis Index (TLI)	≥ .90	.987	.997
Comparative Fit Index (CFI)	≥ .90	.973	.999
Root Mean Square Error of Approximation (RMSEA)	≤ .08	.114	.040
Hoelter's Critical N	≥ 75	145	597

After achieving the model fit, the next step is to evaluate and interpret the importance of indicators in the measurement model. The examination of Table 18 shows that all of the indicators have standardized regression weights ranging from .833 to .934, meaning that they all are important factors influencing the validity of performance measures. According to the table, the most important indicator is *being clear/understandable* (V5) with a standardized regression weight of .934, whereas *focus on importance* (V3) is the least important indicator with a value of .833.

5.3.4.2 Legitimacy. Legitimacy is the second dimension of the latent variable, quality of performance measures. The proposed measurement model for legitimacy has six indicators (see Table 1 and Figure 8). Since none of the indicators had high correlation among each

other, all of the indicators in the proposed model were kept in the generic model, which is shown in Figure 20.

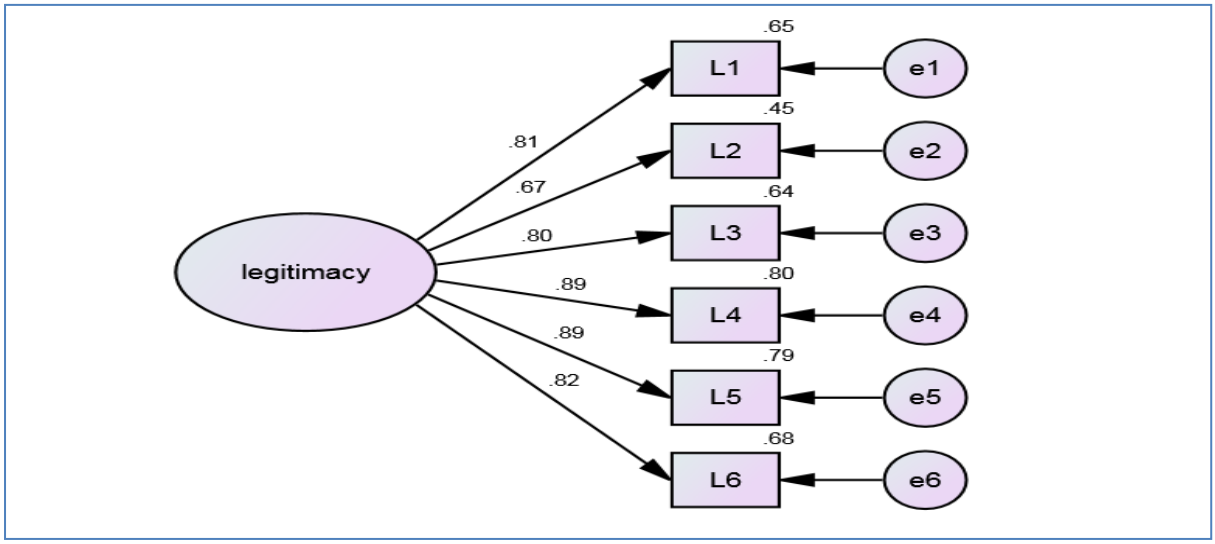

Figure 20. Generic Measurement Model for Legitimacy

First step of the analysis is to examine the significance and strength of factor loadings. Examination of Table 20 shows that all of the indicators of legitimacy had critical ratios higher than 1.96 and p values lower than .05, which means that factor loadings for all indicators show statistical significance. Moreover, all of the factor loadings were found to be important, since they exceeded the predetermined criteria of .30. Therefore, all of the indicators were decided to be kept in the model.

On the other hand, examination of the goodness of fit statistics in Table 21 revealed that some of the indices, like CFI and TLI, showed a good fit, whereas the others were below the threshold levels determined for this study. Therefore, a revision was done in the model to achieve a better fit. For the revision, measurement errors of four pairs of indicators were correlated. The revised measurement model for legitimacy is presented in Figure 21.

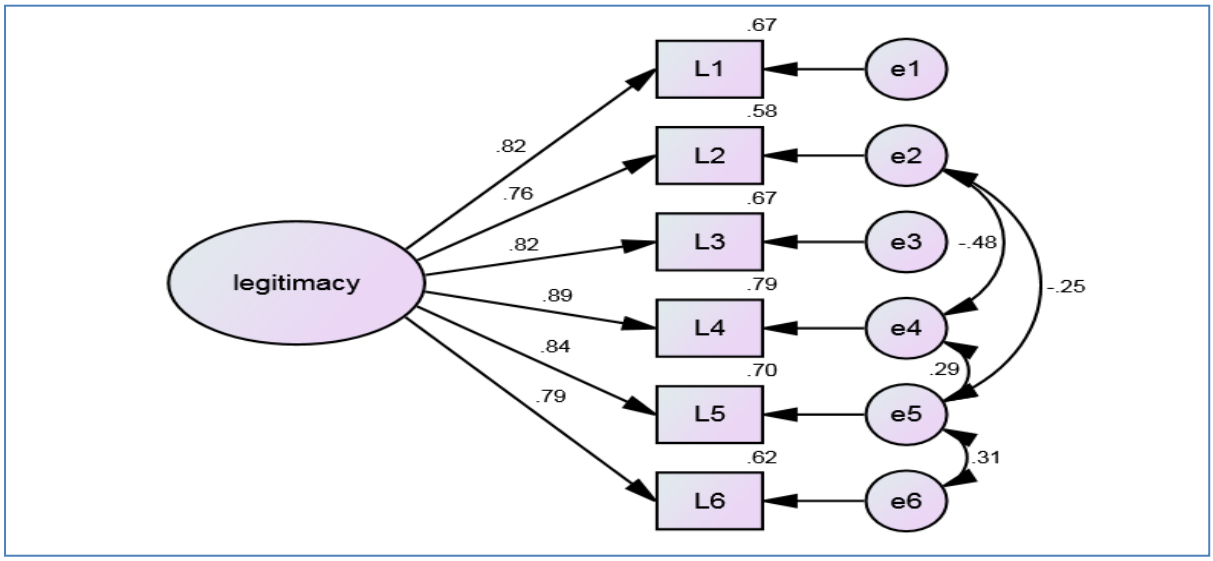

Figure 21. Revised Measurement Model for Legitimacy

Parameter estimates of both the generic model and the revised model are given in Table 20. The examination of the table shows that all of the indicators and the correlations among the measurement errors have statistical significance in both models. Moreover, all of the factor loadings exceeded .30, showing that they are important factors for the legitimacy of performance measures. As a result, all of the indicators were decided to be kept in the revised model.

Table 20 *Parameter Estimates for Legitimacy*

	Generic Model					Revised Model				
Indicator	URW	SRW	SE	CR	P	URW	SRW	SE	CR	P
L1	1.000	.809				1.000	.818			
L2	.796	.668	.053	14.921	***	.897	.760	.053	17.015	***
L3	.980	.799	.052	18.889	***	.996	.820	.050	20.033	***
L4	1.055	.893	.048	22.140	***	1.039	.888	.049	21.413	***
L5	1.016	.887	.046	21.929	***	.946	.836	.049	19.491	***
L6	.944	.822	.048	19.650	***	.894	.786	.048	18.804	***
e2<--> e4						-.187	-.484	.028	-6.698	***
e2<--> e5						.127	.310	.025	5.107	***
e4<--> e5						-.111	-.248	.027	-4.123	***
e5<--> e6						.092	.293	.025	3.680	***

Note: URW = Unstandardized Regression Weight; SRW = Standardized Regression Weight; SE = Standard Error; CR = Critical Ratio; *** = Correlation is significant at .01 level

Goodness of fit statistics of both generic and revised models are given in Table 21. The table indicates that most of the statistics show poor fit for the generic model, but they all show excellent fit for the revised model.

Table 21 *Goodness of Fit Statistics for Legitimacy*

Fit Index	Criteria	Generic Model	Revised Model
Chi-Square (CMIN)	Smaller the better	105.256	6.158
Chi-Square related p value	≥ .05	.000	.291
Chi-square / Degree of Freedom (CMIN /df)	≤ 4	11.695	1.232
Tucker Lewis Index (TLI)	≥ .90	.915	.998
Comparative Fit Index (CFI)	≥ .90	.949	.999
Root Mean Square Error of Approximation (RMSEA)	≤ .08	.158	.023
Hoelter's Critical N	≥ 75	69	768

After achieving the model fit, the next step is to evaluate and interpret the importance of indicators in the measurement model. The examination of Table 20 shows that all of the indicators have standardized regression weights ranging from .760 to .888, meaning that they all are important factors influencing the legitimacy of performance measures. According to the table, the most important indicator is *perceived usefulness by elected officials* (L4) with a standardized regression weight of .888, whereas *involvement of employees* (L2) is the least important indicator with a value of .760.

5.3.4.3 Functionality. The third dimension of the variable quality of performance measures is functionality. The proposed measurement model for functionality has five indicators (see Table 1 and Figure 8). Since none of the indicators had high correlation among each other, all of the indicators in the proposed model were kept in the generic model, which is shown in Figure 22.

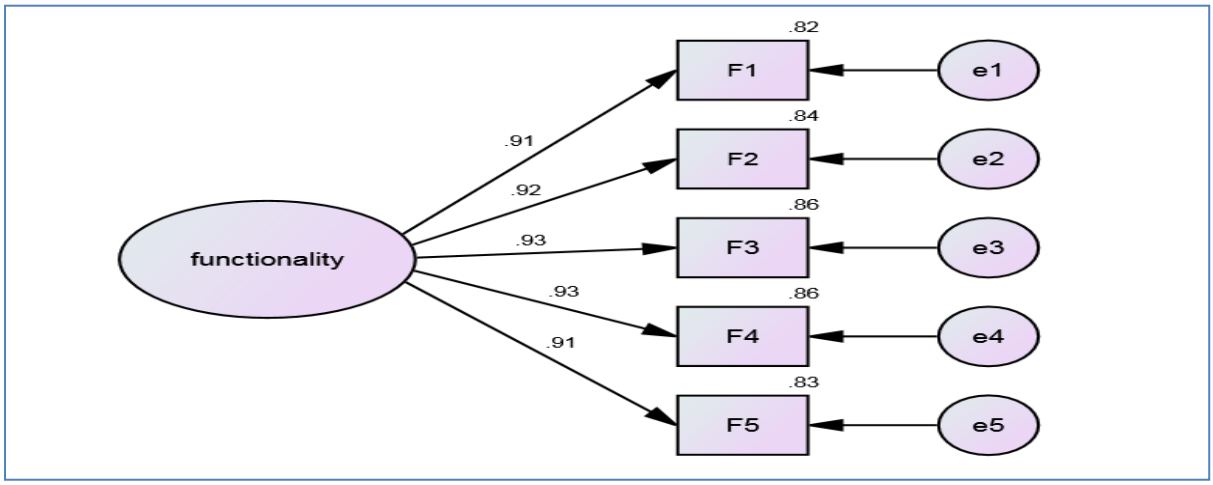

Figure 22. Generic Measurement Model for Functionality

First step of the analysis is to examine the significance and strength of the factor loadings. Examination of Table 22 shows that all of the indicators of functionality in the generic model had critical ratios higher than 1.96 and p values lower than .05, which means that factor loadings for all indicators show statistical significance. Moreover, all of the factor loadings were found to be important, since they exceeded the predetermined criteria of .30. Therefore, all of the indicators were decided to be kept in the model.

On the other hand, examination of the goodness of fit statistics in Table 23 revealed that some of the indices, like CFI and TLI, showed a good fit, whereas the others were below the threshold levels determined for the study. Therefore, a revision was done in the model to achieve a better fit. For the revision, measurement errors of two pairs of indicators were correlated. The revised measurement model for functionality is presented in Figure 23.

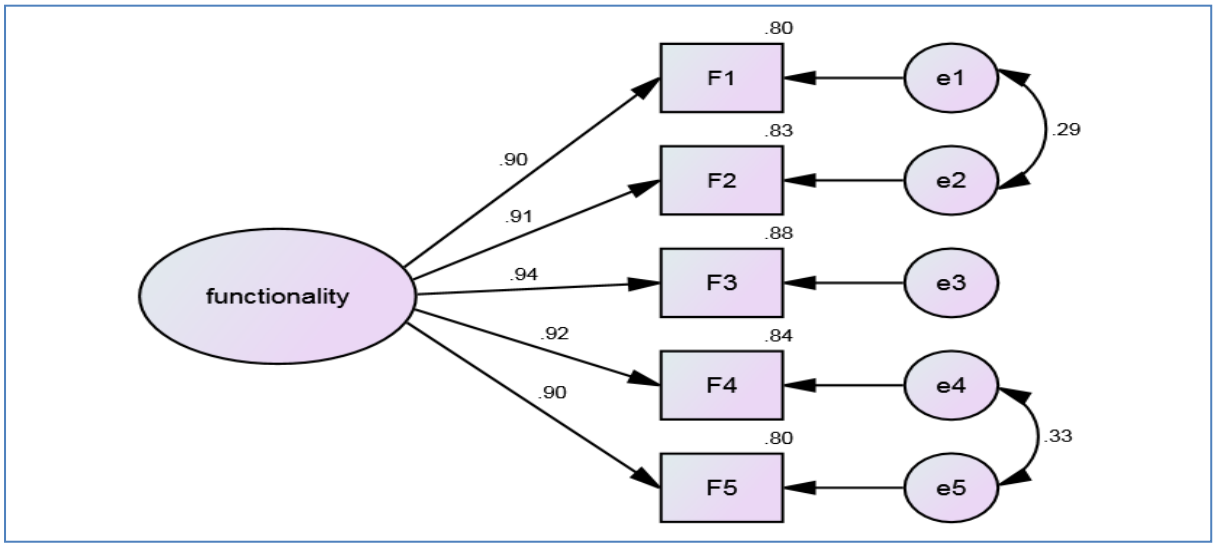

Figure 23. Revised Measurement Model for Functionality

Parameter estimates of both the generic model and the revised model are given in Table 22. The examination of the table shows that all of the indicators and the correlations among the measurement errors have statistical significance. Moreover, all of the factor loadings exceeded .30, showing that they are important factors for the functionality of performance measures. As a result, all of the indicators were decided to be kept in the revised model.

Table 22 *Parameter Estimates for Functionality*

	Generic Model					Revised Model				
Indicator	URW	SRW	SE	CR	P	URW	SRW	SE	CR	P
F1	1.000	.906				1.000	.895			
F2	.983	.916	.031	31.390	***	.988	.910	.028	35.204	***
F3	.992	.930	.030	32.738	***	1.015	.940	.032	31.758	***
F4	1.002	.929	.031	32.707	***	1.002	.919	.034	29.855	***
F5	.985	.911	.032	30.925	***	.981	.897	.035	28.061	***
e1<--> e2						.060	.290	.015	4.031	***
e4<--> e5						.064	.329	.015	4.385	***

Note: URW = Unstandardized Regression Weight; SRW = Standardized Regression Weight; SE = Standard Error; CR = Critical Ratio; *** = Correlation is significant at .01 level

Goodness of fit statistics for both the generic and the revised models are given in Table 23. The table indicates that some of the statistics showed poor fit for the generic model, but they all show excellent fit for the revised model.

Table 23 *Goodness of Fit Statistics for Functionality*

Fit Index	Criteria	Generic Model	Revised Model
Chi-Square (CMIN)	Smaller the better	64.766	3.805
Chi-Square related p value	≥ .05	.000	.283
Chi-square / Degree of Freedom (CMIN /df)	≤ 4	12.953	1.268
Tucker Lewis Index (TLI)	≥ .90	.954	.999
Comparative Fit Index (CFI)	≥ .90	.977	1.000
Root Mean Square Error of Approximation (RMSEA)	≤ .08	.167	.025
Hoelter's Critical N	≥ 75	73	878

After achieving the model fit, the next step is to evaluate and interpret the importance of indicators in the measurement model. The examination of Table 22 shows that all of the indicators have standardized regression weights ranging from .895 to .940, meaning that they all are important factors influencing the functionality of performance measures. According to the table, the most important indicator is *potential for increasing employee motivation* (F3) with a standardized regression weight of .940, whereas *potential for service quality improvement* (F1) is the least important indicator with a value of .895.

5.3.4.4 Integrated Model. After validating the measurement models of three dimensions of quality of performance measures, next step is to integrate these models into one measurement model for quality of performance measures. The integrated measurement

model consists of three first-order variables and a total of 16 indicators. The model is shown in Figure 24.

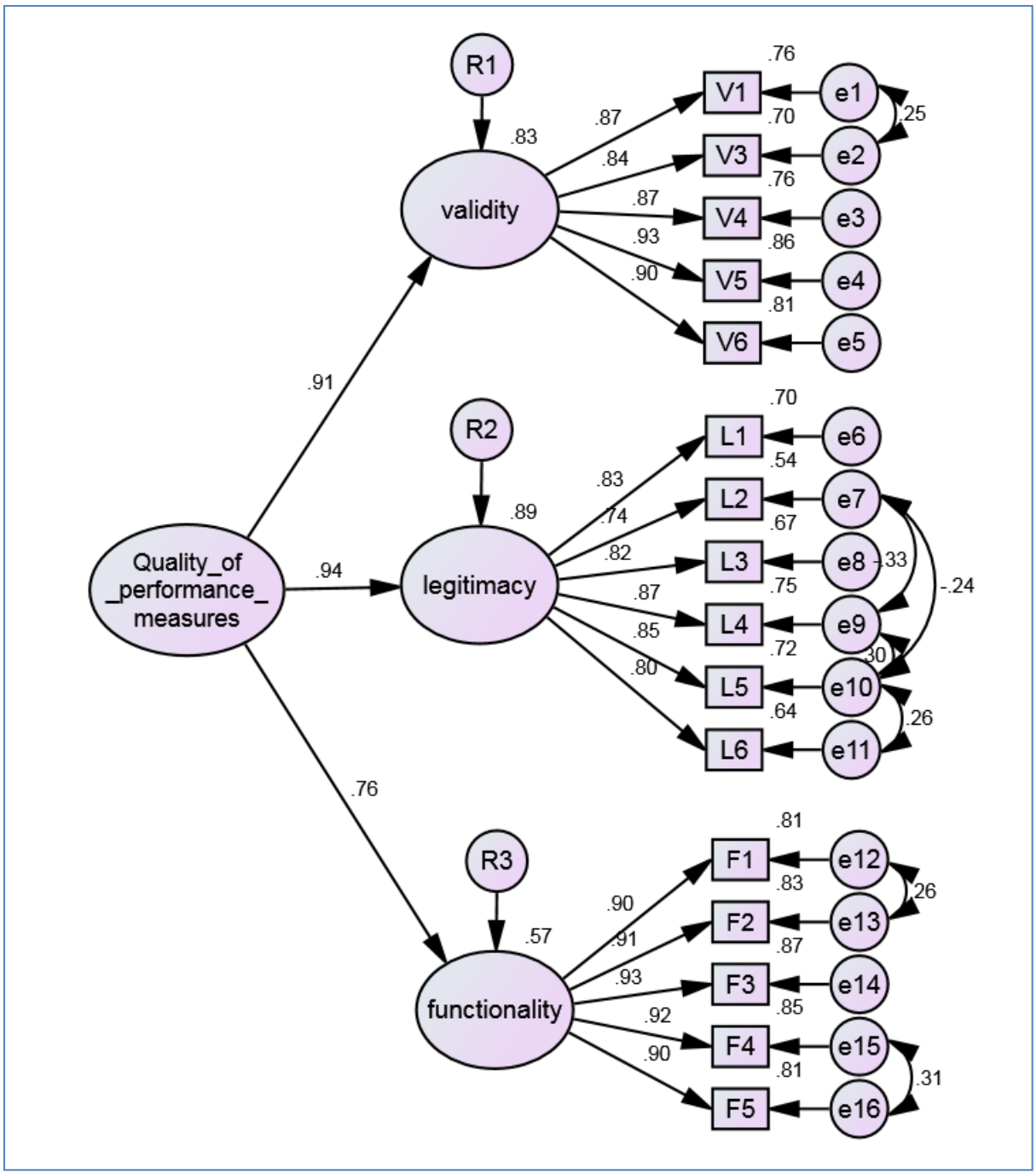

Figure 24. Generic Measurement Model for Quality of Performance Measures

Like the previous models, the first step is to examine the significance and strength of the factor loadings. Parameter estimates of the model are given in Table 24. Examination of

Table 24 shows that all of the indicators of quality of performance measures had critical ratios higher than 1.96 and p values lower than .05, which means that factor loadings for all indicators show statistical significance. Moreover, all of the factor loadings were found to be important, since they exceeded the predetermined criteria of .30. Therefore, all of the indicators were decided to be kept in the model.

Table 24 *Parameter Estimates for Quality of Performance Measures*

	Generic Model				
Indicator	URW	SRW	SE	CR	P
validity<--- QPM	1.000	.910			
legitimacy<--- QPM	.972	.944	.059	16.515	***
functionality<-- QPM	.766	.758	.050	15.360	***
V1	1.000	.869			
V3	.888	.839	.033	26.835	***
V4	1.043	.874	.042	25.061	***
V5	1.022	.929	.036	28.456	***
V5	1.039	.834	.039	26.791	***
L1	1.000	.738			
L2	.854	.820	.049	17.286	***
L3	.976	.902	.047	20.637	***
L4	.995	.912	.045	22.230	***
L5	.945	.934	.044	21.439	***
L6	.891	.920	.045	19.787	***
F1	1.000	.900			
F2	.983	.868	.028	35.402	***
F3	1.002	.851	.031	32.087	***
F4	.997	.903	.033	30.636	***
F5	.978	.799	.034	28.898	***
e1<--> e2	.092	.253	.022	4.181	***
e7<--> e9	-.142	-.326	.026	-5.390	***
e7<--> e10	-.105	-.237	.025	-4.265	***
e9<--> e10	.096	.296	.022	4.409	***
e10<--> e11	.098	.258	.022	4.443	***
e12<--> e13	.053	.263	.014	3.692	***
e15<--> e16	.059	.312	.014	4.246	***

Note: QPM= Quality of Performance Measures; URW = Unstandardized Regression Weight; SRW = Standardized Regression Weight; SE = Standard Error; CR = Critical Ratio; *** = Correlation is significant at .01 level

The next step is to analyze the goodness of fit statistics, which are given in Table 25. Almost all of the indices, except chi-square associated p value, show very good fit of the model to the data at hand. Chi-square associated p has a value of .000, indicating that there is a significant discrepancy between the proposed model and the actual model, meaning a poor model fit.

Table 25 *Goodness of Fit Statistics for Quality of Performance Measures*

Fit Index	Criteria	Generic Model
Chi-Square (CMIN)	Smaller the better	219.150
Chi-Square related p value	≥ .05	.000
Chi-square / Degree of Freedom (CMIN /df)	≤ 4	2.331
Tucker Lewis Index (TLI)	≥ .90	.978
Comparative Fit Index (CFI)	≥ .90	.983
Root Mean Square Error of Approximation (RMSEA)	≤ .08	.056
Hoelter's Critical N	≥ 75	230

However, as mentioned in the methodology section, chi-square is criticized by its sensitivity to sample size (Garson, 2012). In a larger sample size, it is more likely to flag even very trivial differences as significant and to reject something true (type II error) (Ullman, 2007). Therefore, Garson (2012) and Schermelleh-Engel et al. (2003) argue that for many researchers finding significant chi-square is not a reason for model modification, if sample size is over 200 and other fit indices show a good fit. Since other indices, including relative chi-square (CMIN/df), show a very good fit and the sample size (428) of the study is highly over 200, a modification of the model was not seen as necessary.

After achieving the model fit, the next step is to evaluate and interpret the importance of indicators in the measurement model. The examination of Table 24 shows that all of the indicators have standardized regression weights ranging from .738 to .934, meaning that they all are important factors influencing their respective first-order variables. The regression weights of the three dimensions, validity, legitimacy, and functionality, also show that they are important for the second-order variable, quality of performance measures. Based on the analysis, legitimacy seems to be the most important dimension with a standardized regression weight of .944. On the other hand, functionality has only a regression weight of .758, making this dimension as the least important for the indicating the level of quality of performance measures in this study.

5.3.5 Effectiveness of Performance Measurement

Effectiveness of performance measurement is the last variable, for which a measurement model is created. The proposed measurement model had ten indicators (see Table 1 and Figure 9). However, correlation analysis revealed that there is high correlation between three pairs of indicators, which are E1-E2, E2-E4, and E7-E8. Moreover, another pair (E8-E9) has a correlation value (.848) which is very close to the threshold level (.85). In order to eliminate the high risk of multicollinearity, two indicators, namely *improvement in service quality* (E2) and *better solution of problems* (E8) were excluded from the model. As a result, generic model for effectiveness of performance measurement consisted of eight indicators. The model is presented in Figure 25.

Similar to the other models, the first step is to examine the significance and strength of the factor loadings. Examination of Table 26 shows that all of the indicators of in the generic model had critical ratios higher than 1.96 and p values lower than .05, which means that factor loadings for all indicators show statistical significance. Moreover, all of the factor

loadings were found to be important, since they exceeded the .30 level. Therefore, all of the indicators were decided to be kept in the model.

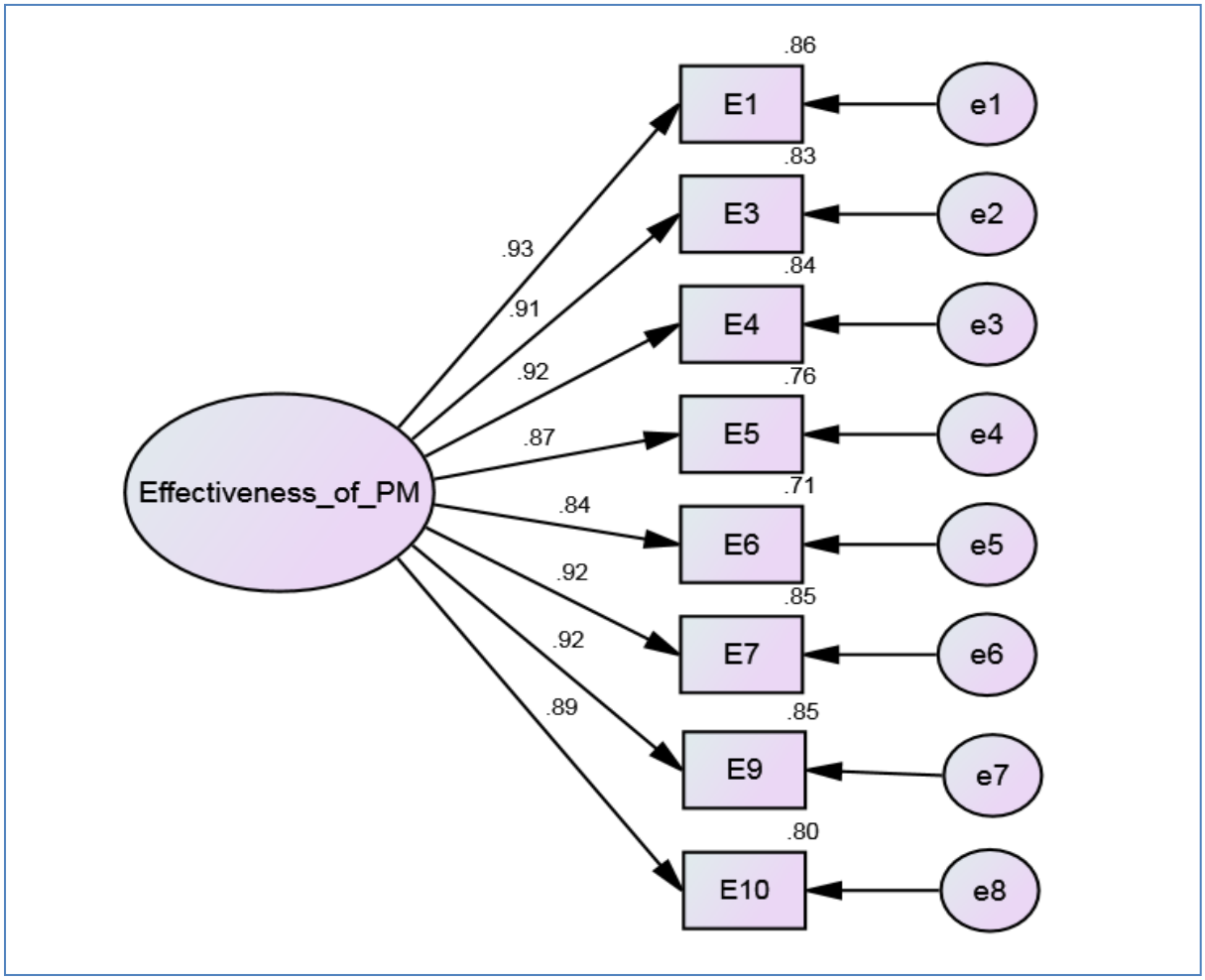

Figure 25. Generic Measurement Model for Effectiveness of Performance Measurement

On the other hand, examination of the goodness of fit statistics in Table 27 revealed that some of the indices, like CFI, TLI, and Hoelter's Critical N, showed a good fit, whereas the others were below the threshold levels determined for the study. Therefore, a revision was done in the model to achieve a better fit. For the revision, measurement errors of seven pairs of indicators were correlated. The revised measurement model for effectiveness of performance measurement is presented in Figure 26.

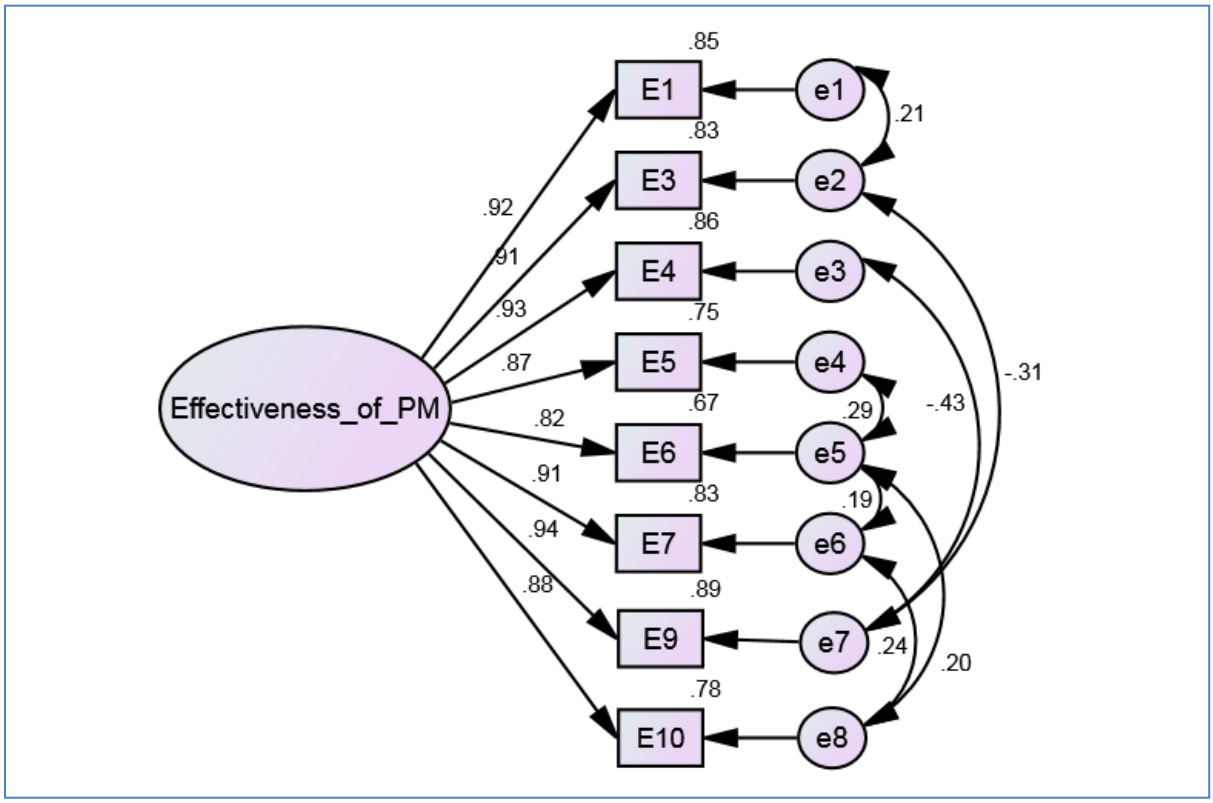

Figure 26. Revised Measurement Model for Effectiveness of Performance Measurement

Parameter estimates of both the generic model and the revised model are given in Table 26. Examination of the table shows that all of the indicators and the correlations among the measurement errors in the revised model have statistical significance. Moreover, all of the factor loadings exceeded .30, showing that they are important factors for the effectiveness of performance measurement. As a result, all of the indicators were decided to be kept in the revised model.

Goodness of fit statistics for both the generic and revised models are given in Table 27. The table indicates that some of the statistics showed poor fit for the generic model, but they all show excellent fit for the revised model, which means that the revised measurement model fits the data at hand. As a result, it can be concluded that measurement is validated.

Table 26 *Parameter Estimates for Effectiveness of Performance Measurement*

	Generic Model					Revised Model				
Indicator	URW	SRW	SE	CR	P	URW	SRW	SE	CR	P
E1	1.000	.929				1.000	.923			
E3	.972	.909	.029	33.407	***	.982	.912	.027	36.831	***
E4	.968	.916	.028	34.265	***	.989	.929	.029	34.599	***
E5	.932	.873	.031	29.602	***	.932	.866	.033	28.611	***
E6	.928	.843	.034	27.044	***	.906	.817	.036	24.867	***
E7	.966	.924	.027	35.353	***	.962	.913	.029	33.307	***
E9	.951	.921	.027	34.919	***	.981	.942	.027	36.185	***
E10	.934	.894	.029	31.704	***	.929	.882	.031	30.073	***
e1<--> e2						.035	.206	.012	3.028	.002
e2<--> e7						-.044	-.310	.010	-4.423	***
e3<--> e7						-.055	-.435	.009	-5.887	***
e4<--> e5						.091	.287	.016	5.577	***
e5<--> e6						.047	.187	.013	3.628	***
e5<--> e8						.057	.197	.015	3.895	***
e6<--> e8						.048	.244	.011	4.202	***

Note: URW = Unstandardized Regression Weight; SRW = Standardized Regression Weight; SE = Standard Error; CR = Critical Ratio; *** = Correlation is significant at .01 level

Table 27 *Goodness of Fit Statistics for Effectiveness of Performance Measurement*

Fit Index	Criteria	Generic Model	Revised Model
Chi-Square (CMIN)	Smaller the better	127.653	17.992
Chi-Square related p value	≥ .05	.000	.158
Chi-square / Degree of Freedom (CMIN /df)	≤ 4	6.383	1.384
Tucker Lewis Index (TLI)	≥ .90	.966	.998
Comparative Fit Index (CFI)	≥ .90	.975	.999
Root Mean Square Error of Approximation (RMSEA)	≤ .08	.112	.030
Hoelter's Critical N	≥ 75	106	531

After achieving the model fit, the next step is to evaluate and interpret the importance of indicators in the measurement model. Examination of Table 26 shows that all of the indicators have standardized regression weights ranging from .817 to .942, meaning that they all are important factors influencing the effectiveness of performance measurement. According to the table, the most important indicator is *better decision-making* (E9) with a standardized regression weight of .942, whereas *cost reduction* (E6) is the least important indicator with a value of .817.

In this section, the process of validating the measurement models of the study through the confirmatory factor analysis was illustrated and explained. Since all of the measurement models are validated, they are ready for further analysis of reliability (internal consistency) and later for structural equation modeling.

5.4 Reliability Analysis

Measurement reliability is an important part of any study which includes scale variables. Measurement reliability concerns on non-systematic, or random errors. It mainly examines whether measurement creates consistent results over time (Babbie, 2013). Since the scales for latent variables in the study were not taken from the literature as a whole, it is also important to test the measurement reliability of the scales used.

Cronbach's alpha score, which is one of the most common methods in measuring the reliability (Kline, 2011; Streib & Poister, 1999), was used in the study. The analysis was run by SPSS program. In the literature, it is argued that if the alpha score exceeds the generally accepted adequate level of .70 (de Lancer Julnes & Holzer, 2001; George & Mallery, 2007; Kline, 2011; Morgan et al., 2005), it means that the measurement produces consistent results at different times (Cronbach, 1951). An alpha score greater than .80 is considered as good

and greater than .90 is considered as excellent (George & Mallery, 2007). In this study, .70 was set as the threshold for Cronbach's alpha.

Table 28 presents Cronbach's alpha values for each latent constructs before and after the data analysis. The α values ranged from .890 to .979 before conducting any analysis. A total of five indicators were removed from the model based on the results of correlation analysis and confirmatory factor analysis. Findings of the reliability analysis after the data analyses still showed almost excellent reliability, ranging from .882 to .972. The scale of quality of performance measures had the highest alpha scores both before and after the data analyses with .979 and .972 respectively. On the other hand, scale of external support received the lowest scores in both times with .890 and .882. In general, the results show that the scales used in the study has almost excellent measurement reliability, meaning that they produce consistent results at different times.

Table 28 *Cronbach's Alpha Values for the Measurement Models*

Latent Construct	Number of Items		Cronbach's Alpha (α)	
	Before	After	Before	After
Organizational Support	7	7	.900	.900
External Support	5	4	.890	.882
Technical Capacity	6	5	.942	.924
Validity of Performance Measures	6	5	.958	.947
Legitimacy of Performance Measures	6	6	.922	.922
Functionality of Performance Measures	5	5	.964	.964
Effectiveness of Performance Measurement	10	8	.979	.972

Since the measurement models were validated through CFA and almost all of the scales showed excellent reliability, the next step is to combine these measurement models in a covariance structure model, so that the hypotheses of the study can be tested through structural equation modeling.

5.5 Structural Equation Modeling

The last stage of the statistical analysis is structural equation modeling (SEM). It is an analytical method used to test the hypothesized relationships among the study variables. Validated measurement models explained in the previous parts were combined into one model, namely covariance structure model. This model also includes the control variables of the study.

The analysis was carried out mainly according to the three-step method proposed by Wan (2002). These steps include; 1) checking the appropriateness of the indicators, 2) checking the overall model fit, 3) revising the model. However, in the first step, not only the significance of indicators, but also that of path coefficients was examined. Path coefficients show the strength of relationship between the study variables, including control variables. The generic covariance structure model is presented in Figure 27.

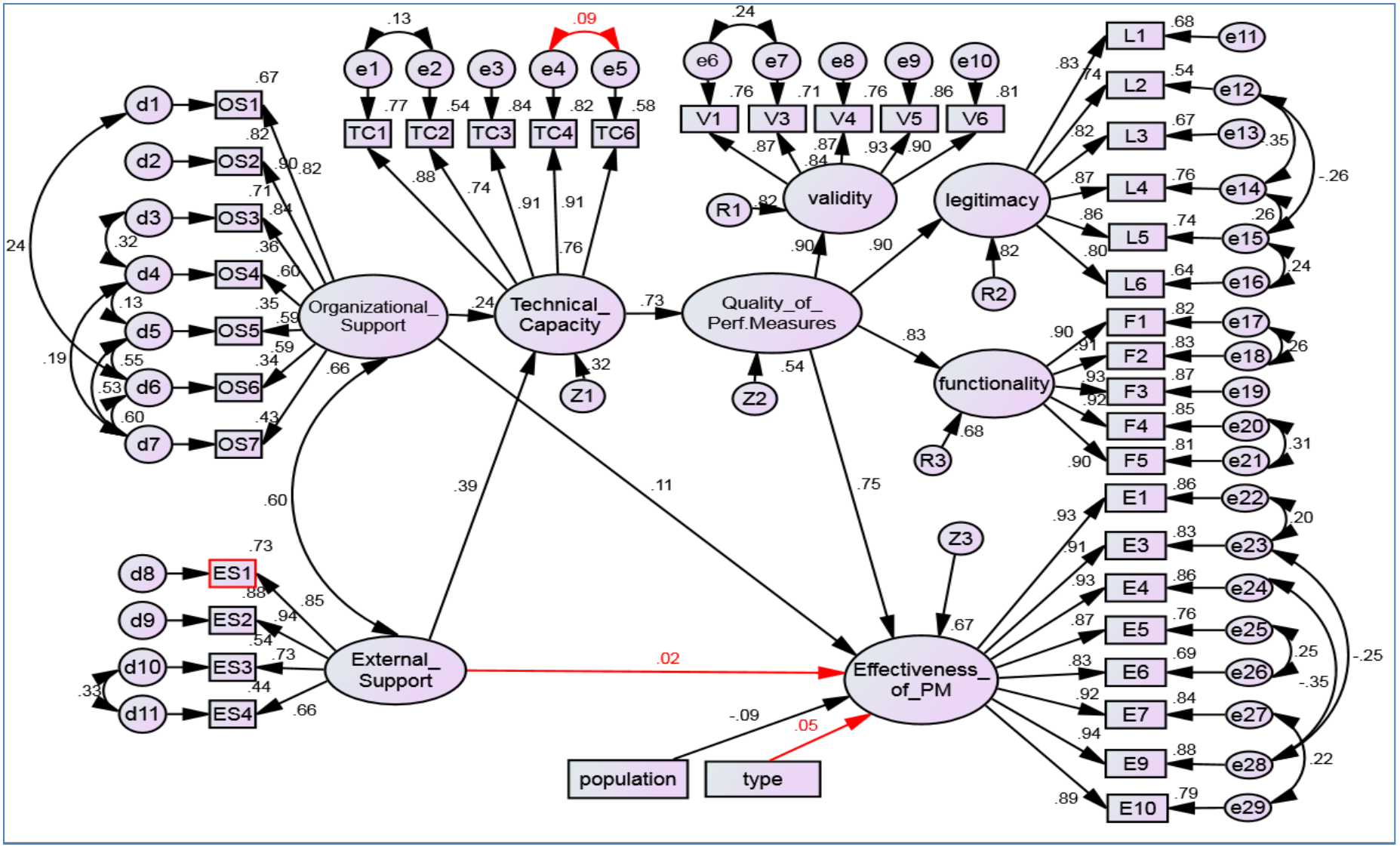

Figure 27. Generic Covariance Structure Model

Similar to the analysis of the measurement models, firstly, the parameter estimates of the study were examined. Examination of Table 29 shows that all of the indicators of in the generic model had critical ratios higher than 1.96 and p values lower than .05, which means that factor loadings for all indicators show statistical significance. Moreover, all of the factor loadings were found to be important, since they exceeded the .30 level. Examination also revealed that all but one of the correlations among the measurement errors were statistically significant. The correlation among the measurement errors of two indicators of technical capacity, namely TC4 and TC6, failed to achieve the significance level in the combined model, although it was significant in the measurement model of technical capacity. Therefore, it was excluded from the model.

Lastly, the significance of path coefficients among the study variables was controlled. Two path coefficients were found to be statistically insignificant. The first one is the path between the external support and effectiveness of performance measurement. The second one is the path between the control variable, type, and effectiveness of performance measurement. As a result, these paths were removed from the model. All statistically insignificant relationships are shown in red in Figure 27.

The examination of the goodness of fit statistics in Table 30 also revealed that almost all of the indices showed an acceptable fit. Only chi-square associated p value did not show an acceptable fit. Although modification of a model is not required only because of model chi-square, in order to exclude the above-mentioned insignificant relationships and to have a better model fit, a revision was conducted. For the revision, other than the exclusion of insignificant relationships, two more pairs of measurement errors, which are e19-e23 and e20-e24, were correlated by using the modification indices. The revised covariance structure model is presented in Figure 28.

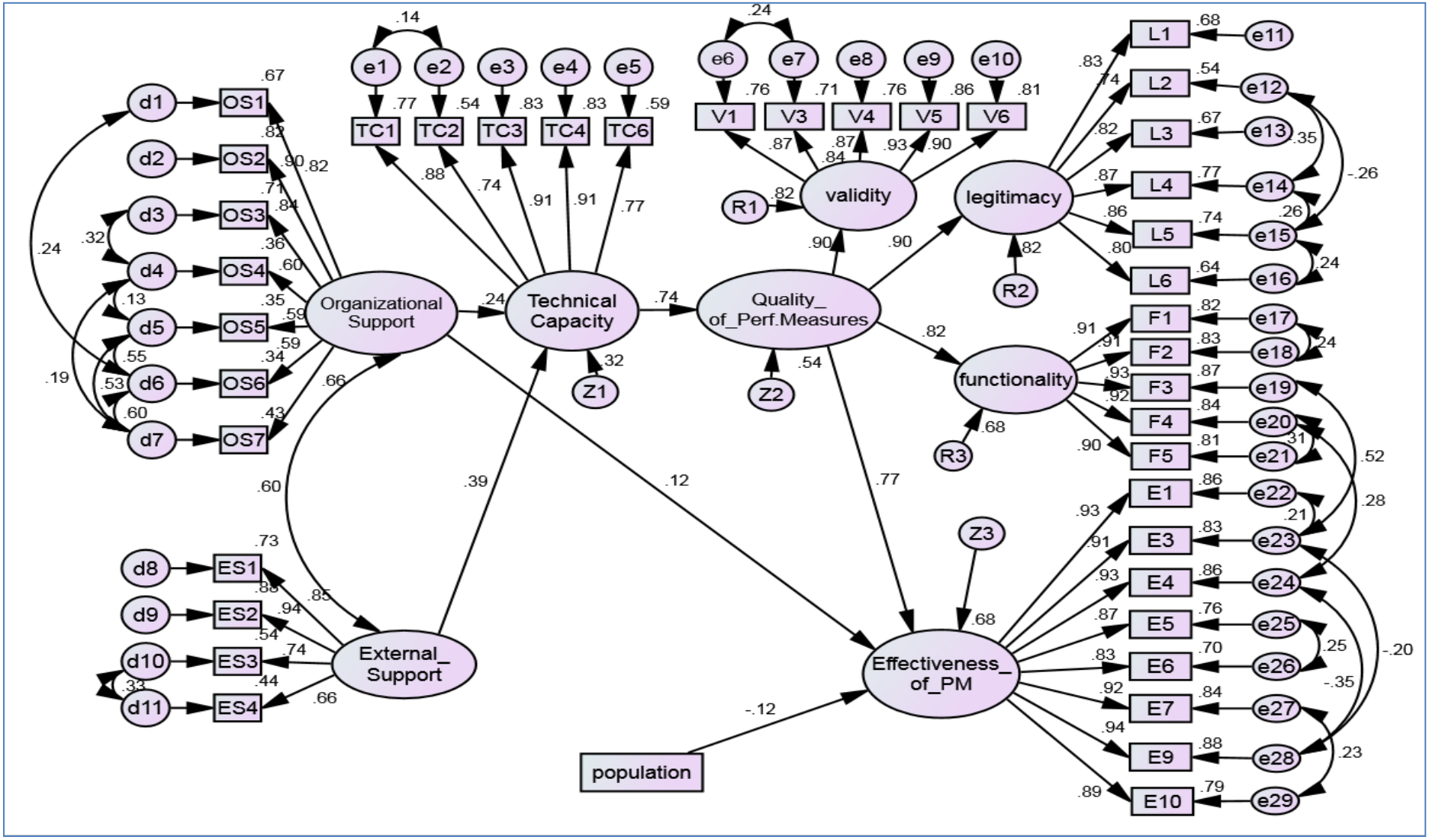

Figure 28. Revised Covariance Structure Model

Table 29 *Parameter Estimates for the Covariance Structure Model*

		Generic Model					Revised Model				
Indicator		URW	SRW	SE	CR	P	URW	SRW	SE	CR	P
Technical Capacity	<--- Organizational Support	.404	.243	.097	4.146	***	.402	.242	.097	4.134	***
Technical Capacity	<--- External Support	.500	.388	.076	6.561	***	.498	.388	.076	6.558	***
Quality of performance measures	<--- Technical Capacity	.610	.734	.041	14.755	***	.614	.736	.042	14.801	***
Effectiveness of PM	<--- Quality of performance measures	.766	.753	.048	15.824	***	.787	.767	.048	16.470	***
validity	<--- Quality of performance measures	1.000	.904				1.000	.904			
legitimacy	<--- Quality of performance measures	.926	.903	.054	17.131	***	.925	.903	.054	17.138	***
functionality	<--- Quality of performance measures	.840	.826	.049	17.228	***	.843	.825	.049	17.331	***
Effectiveness of PM	<--- Organizational Support	.159	.113	.061	2.629	.009	.167	.117	.050	3.358	***
Effectiveness of PM	<--- Population	-.062	-.090	.021	-2.914	.004	-.084	-.120	.021	-3.934	***
Effectiveness of PM	<--- Type	.048	.050	.030	1.616	.106	-	-	-	-	-
Effectiveness of PM	<--- External Support	.027	.024	.047	.561	.575	-	-	-	-	-
OS1	<--- Organizational Support	1.000	.821				1.000	.821			
OS2	<--- Organizational Support	1.250	.905	.057	22.079	***	1.249	.905	.057	22.070	***
OS3	<--- Organizational Support	1.109	.840	.055	20.181	***	1.109	.840	.055	20.176	***
OS4	<--- Organizational Support	.882	.603	.068	12.944	***	.882	.604	.068	12.946	***
OS5	<--- Organizational Support	.978	.589	.077	12.709	***	.979	.590	.077	12.721	***
OS6	<--- Organizational Support	.974	.585	.070	13.959	***	.974	.585	.070	13.968	***
OS7	<--- Organizational Support	1.023	.656	.071	14.490	***	1.024	.656	.071	14.501	***
ES1	<--- External Support	1.000	.852				1.000	.852			
ES2	<--- External Support	1.131	.938	.047	24.036	***	1.131	.938	.047	24.018	***
ES3	<--- External Support	.931	.735	.053	17.626	***	.932	.735	.053	17.637	***
ES4	<--- External Support	.863	.660	.057	15.165	***	.864	.661	.057	15.175	***
TC1	<--- Technical Capacity	1.000	.878				1.000	.875			
TC2	<--- Technical Capacity	.801	.738	.040	19.786	***	.801	.736	.041	19.762	***
TC3	<--- Technical Capacity	1.020	.915	.037	27.646	***	1.021	.913	.037	27.409	***

		Generic Model					Revised Model				
Indicator		URW	SRW	SE	CR	P	URW	SRW	SE	CR	P
TC4	<--- Technical Capacity	1.017	.905	.038	26.942	***	1.025	.910	.038	27.239	***
TC6	<--- Technical Capacity	.871	.762	.045	19.379	***	.884	.771	.044	19.976	***
V1	<--- validity	1.000	.871				1.000	.871			
V3	<--- validity	.889	.842	.033	26.918	***	.889	.842	.033	26.917	***
V4	<--- validity	1.039	.873	.041	25.158	***	1.039	.873	.041	25.148	***
V5	<--- validity	1.017	.927	.036	28.531	***	1.017	.927	.036	28.522	***
V6	<--- validity	1.036	.902	.038	26.947	***	1.037	.903	.038	26.956	***
L1	<--- legitimacy	1.000	.827				1.000	.827			
L2	<--- legitimacy	.862	.738	.050	17.094	***	.862	.738	.050	17.100	***
L3	<--- legitimacy	.980	.816	.048	20.258	***	.980	.816	.048	20.260	***
L4	<--- legitimacy	1.011	.875	.046	22.137	***	1.012	.875	.046	22.148	***
L5	<--- legitimacy	.963	.860	.045	21.439	***	.963	.860	.045	21.441	***
L6	<--- legitimacy	.902	.802	.046	19.669	***	.902	.802	.046	19.658	***
F1	<--- functionality	1.000	.903				1.000	.908			
F2	<--- functionality	.981	.912	.028	35.466	***	.976	.912	.027	35.679	***
F3	<--- functionality	.999	.933	.031	32.195	***	.990	.932	.029	33.681	***
F4	<--- functionality	.995	.920	.032	30.840	***	.986	.918	.031	31.687	***
F5	<--- functionality	.977	.901	.034	29.137	***	.970	.899	.033	29.648	***
E1	<--- Effectiveness of PM	1.000	.928				1.000	.928			
E3	<--- Effectiveness of PM	.979	.913	.026	37.559	***	.970	.911	.025	39.179	***
E4	<--- Effectiveness of PM	.984	.928	.028	35.375	***	.984	.930	.027	36.126	***
E5	<--- Effectiveness of PM	.930	.870	.032	29.278	***	.931	.872	.031	29.561	***
E6	<--- Effectiveness of PM	.915	.831	.035	26.116	***	.916	.834	.035	26.394	***
E7	<--- Effectiveness of PM	.961	.917	.028	34.242	***	.958	.915	.028	34.185	***
E9	<--- Effectiveness of PM	.971	.938	.027	36.606	***	.971	.939	.026	36.905	***
E10	<--- Effectiveness of PM	.930	.888	.030	31.051	***	.930	.890	.030	31.331	***
External Support	<--> Organizational Support	.367	.595	.041	9.035	***	.368	.596	.041	9.042	***
d6	<--> d7	.454	.598	.043	10.480	***	.454	.598	.043	10.475	***

Indicator		Generic Model					Revised Model				
		URW	SRW	SE	CR	P	URW	SRW	SE	CR	P
d6	<--> d5	.474	.549	.048	9.842	***	.474	.549	.048	9.838	***
d5	<--> d4	.097	.130	.030	3.278	.001	.097	.129	.029	3.274	.001
d4	<--> d3	.128	.321	.023	5.461	***	.128	.321	.023	5.460	***
d5	<--> d7	.399	.529	.044	9.099	***	.398	.528	.044	9.090	***
d4	<--> d7	.128	.195	.026	4.971	***	.127	.194	.026	4.967	***
d6	<--> d1	.106	.237	.019	5.591	***	.106	.237	.019	5.590	***
e17	<--> e18	.051	.256	.014	3.639	***	.046	.236	.013	3.447	***
e20	<--> e21	.058	.309	.014	4.261	***	.059	.308	.013	4.594	***
e1	<--> e2	.066	.125	.032	2.089	.037	.073	.136	.032	2.300	.021
e4	<--> e5	.041	.087	.030	1.343	.179	-	-	-	-	-
e6	<--> e7	.086	.240	.022	3.960	***	.086	.240	.022	3.967	***
e22	<--> e23	.033	.199	.011	3.015	.003	.036	.212	.010	3.568	***
e23	<--> e28	-.038	-.254	.010	-3.960	***	-.030	-.200	.009	-3.472	***
e24	<--> e28	-.047	-.348	.009	-5.197	***	-.047	-.352	.009	-5.388	***
e25	<--> e26	.077	.251	.016	4.668	***	.076	.249	.016	4.621	***
e27	<--> e29	.042	.222	.011	3.869	***	.045	.231	.011	4.043	***
e12	<--> e14	-.148	-.348	.026	-5.652	***	-.148	-.350	.026	-5.671	***
e12	<--> e15	-.114	-.263	.025	-4.612	***	-.114	-.263	.025	-4.621	***
e14	<--> e15	.081	.264	.021	3.809	***	.081	.264	.021	3.798	***
e15	<--> e16	.089	.241	.022	4.053	***	.089	.242	.022	4.067	***
d11	<--> d10	.224	.333	.039	5.767	***	.223	.332	.039	5.749	***
e23	<--> e19						.084	.517	.010	8.097	***
e24	<--> e20						.045	.284	.009	4.947	***

Note: URW = Unstandardized Regression Weight; SRW = Standardized Regression Weight; SE = Standard Error; CR = Critical Ratio; *** = Correlation is significant at .01 level

Parameter estimates of both the generic model and the revised model are given in Table 29. Examination of the table shows that all of the indicators and the correlations among the measurement errors in the revised model have statistical significance. Moreover, all of the factor loadings exceeded .30, showing that they are important factors for their respective latent variables. The path coefficients between the study variables showed also statistical significance. As a result, there are not any other paths to be removed from the model.

Goodness of fit statistics for both the generic and revised models are given in Table 30. The statistics show that even the generic model has acceptable fit. Only chi-square associated p value was statistically significant, indicating that there is a significant discrepancy between the proposed model and the actual model. The values for the revised model show a slightly better fit compared to the generic model. But, still, chi-square associated p value is statistically significant.

Table 30 *Goodness of Fit Statistics for the Covariance Structure Model*

Fit Index	Criteria	Generic Model	Revised Model
Chi-Square (CMIN)	Smaller the better	2440.504	2119.775
Chi-Square related p value	$\geq .05$	.000	.000
Chi-square / Degree of Freedom (CMIN /df)	≤ 4	3.101	2.838
Tucker Lewis Index (TLI)	$\geq .90$	.904	.919
Comparative Fit Index (CFI)	$\geq .90$	.912	.927
Root Mean Square Error of Approximation (RMSEA)	$\leq .08$	.070	.066
Hoelter's Critical N	≥ 75	150	164

However, as mentioned earlier, chi-square is criticized by its sensitivity to sample size (Garson, 2012). In a larger sample size, it is more likely to flag even very trivial differences as significant (Ullman, 2007). Therefore, Garson (2012) and Schermelleh-Engel et al. (2003) argue that for many researchers finding significant chi-square is not a reason for model modification, if sample size is over 200 and other fit indices show a good fit. Since other indices, including relative chi-square (CMIN/df), show a good fit and the sample size of the study is highly over 200, a modification of the model was not seen as necessary.

After achieving the model fit, the next step is to evaluate and interpret the importance of study variables in the covariance structure model. Examination of Table 29 shows that all of the paths in the revised model are statistically significant, meaning that the hypothesized relationships between these variables existed in fact according to the data at hand.

The results of the SEM show that, as hypothesized in the study, both organizational support and external support have an indirect effect over effectiveness of performance measurement via technical capacity and quality of performance measures. Quality of performance measures has the strongest relation with the effectiveness of performance measurement with a standardized regression weight of .767. Moreover, it mediates the relationship between technical capacity and effectiveness of performance measurement. The relationship between technical capacity and quality of performance measures is also very strong, with a standardized regression weight of .736. These two variables mediate the relationships of organizational and external support with effectiveness of performance measurement. Similarly, organizational support and external support are positively related to technical capacity with regression weights of respectively .242 and .388. The indirect relationships of these variables with effectiveness of performance measurement can be calculated by multiplying their regression coefficients with technical capacity to that of

technical capacity and quality of performance measures, and quality of performance measures and effectiveness of performance measurement. The indirect effect of organizational support on quality of performance measures is .137 (.242*.736*.767). The indirect effect of external support on effectiveness of performance measurement is .219 (.388*.736*.767).

Other than these mediated relationships, organizational support is also positively and directly related to effectiveness of performance measurement. The results also indicate that there is a positive and significant correlation among organizational support and external support with a standardized regression weight of .596.

Among the two control variables only population has a significant but negative relationship with effectiveness of performance measurement with a regression weight of -.120. A negative relationship in this occasion means that effectiveness of performance measurement decreases in more populated municipalities.

According to the results of the SEM analysis, organizational support and external support account for 32% of variance in technical capacity. The variation in the quality of performance measures explained by these three variables is 54%. Lastly, all the exogenous, endogenous mediating and the control variable of the study account for 68% of variation in effectiveness of performance measurement, which is a very high percentage.

5.6 Hypothesis Testing

In the study, there are seven hypotheses about the relationships of the study variables. In this part, whether or to what extent these hypotheses are supported will be discussed based on the results of SEM analysis. Standardized regression weights of the relationships between the study variables are shown in the conceptual model of the study, which is a simplified

version of the covariance structure model. The model is presented in Figure 28. The relationships which are found to be insignificant are shown in red.

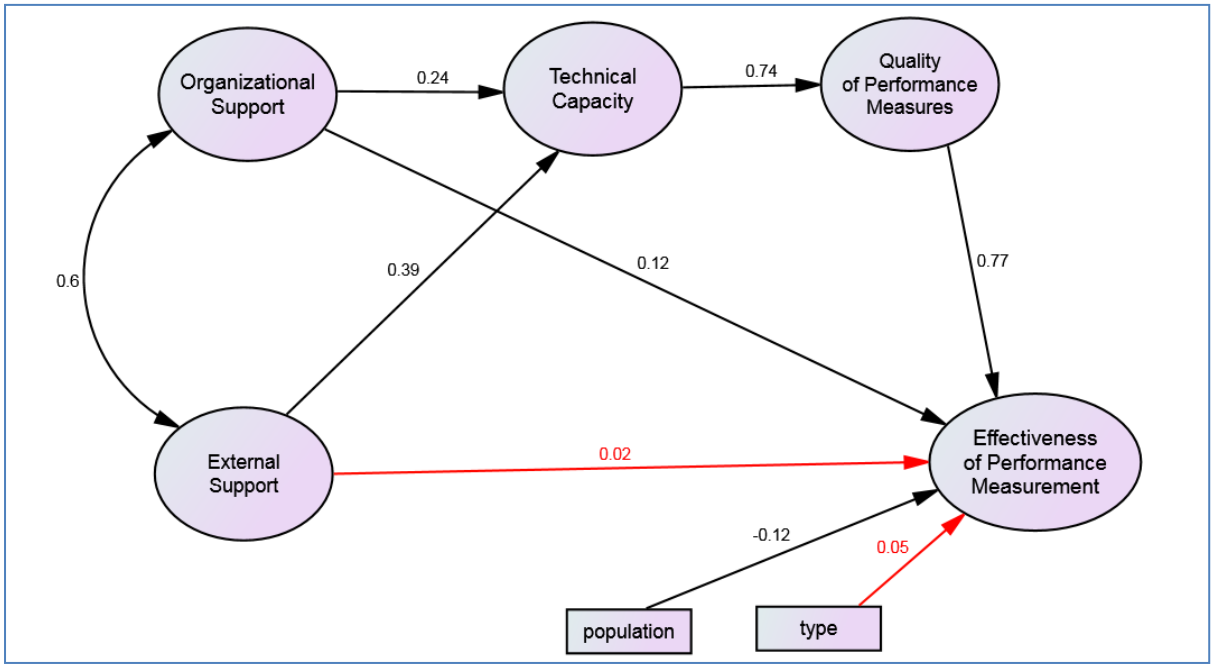

Figure 29. Conceptual Model of the Study with Regression Coefficients

H_1: Quality of performance measures is positively related with the effectiveness of performance measurement systems.

First hypothesis is about relationship between quality of performance measures and effective performance measurement. The results of the analysis show that quality of performance measures is positively and significantly related with effectiveness of performance measurement at .95 confidence level with p<.01, which means that the hypothesis is supported by the data at hand. Unstandardized regression weight of .787 among the variables shows that one unit of increase in quality of performance measures would lead to a .787 unit of increase in effectiveness of performance measurement in Turkish municipalities.

H_2: Technical capacity of the municipality influences the effectiveness of performance measurement indirectly via the quality of performance measures.

In order this hypothesis to be supported, both the relationships between 1) technical capacity and quality of performance measures and 2) quality of performance measures and effectiveness of performance measurement should be statistically significant. As mentioned in the first hypothesis, the second relationship was already significant. The results show that there is also a statistically significant positive relationship between technical capacity and quality of performance measures in Turkish municipalities with the p value lower than .01. As a result, both relationships are statistically significant and the hypothesis is supported.

H_3: External support for performance measurement influences the effectiveness of performance measurement indirectly via technical capacity of the municipality and the quality of performance measures.

In order this hypothesis to be supported, the relationships between 1) external support and technical capacity, 2) technical capacity and quality of performance measures, and 3) quality of performance measures and effectiveness of performance measurement should be statistically significant. As mentioned in the first two hypotheses, last two relationships were already significant. The results show that there is also a statistically significant positive relationship between external support and technical capacity with the p value lower than .001. As a result, all three relationships are statistically significant and the hypothesis is supported.

H_4: External support for performance measurement directly influences the effectiveness of performance measurement.

The results show that the direct relationship of external support with effectiveness of performance measurement is not statistically significant, since the p value is .575, which is higher than the predetermined alpha level of .05. The standardized regression weight of .024 indicates that one standard deviation of variance in external support would lead to an only 2.4% increase in effectiveness of performance measurement, which is very trivial. As a result, this hypothesis is not supported according to the data at hand.

H_5: Organizational support influences the effectiveness of performance measurement indirectly via technical capacity of the municipality and the quality of performance measures.

In order this hypothesis to be supported, the relationships between 1) organizational support and technical capacity, 2) technical capacity and quality of performance measures, and 3) quality of performance measures and effectiveness of performance measurement should be statistically significant. As mentioned in the first two hypotheses, last two relationships were already significant. The results show that there is also a statistically significant relationship between organizational support and technical capacity with the p value lower than .01. As a result, all three relationships are statistically significant and the hypothesis is supported.

H_6: Organizational support directly influences the effectiveness of performance measurement.

The study found that the direct relationship between organizational support and effectiveness of performance measurement is also statistically significant with a p value lower than .01. The standardized regression weight of .117 shows that one standard deviation

of variance in organizational support would lead to a .117 increase in effectiveness of performance measurement in Turkish municipalities. As a result, this hypothesis is supported.

H_7: There is a positive correlation between organizational support and external support.

The results also indicate that there is a positive and significant correlation among organizational support and external support with a standardized regression weight of .596 and with a p value lower than .01. As a result, this hypothesis is supported, as well. The summary of the testing results of the study hypotheses are given in Table 31.

Table 31 *Summary of Hypothesis Testing Results*

	Hypotheses	**Results**
H_1	*Quality of performance measures is positively related with the effectiveness of performance measurement systems.*	Supported
H_2	*Technical capacity of the municipality influences the effectiveness of performance measurement indirectly via the quality of performance measures.*	Supported
H_3	*External support for performance measurement influences the effectiveness of performance measurement indirectly via technical capacity of the municipality and the quality of performance measures.*	Supported
H_4	*External support for performance measurement directly influences the effectiveness of performance measurement.*	Not Supported
H_5	*Organizational support influences the effectiveness of performance measurement indirectly via technical capacity of the municipality and the quality of performance measures.*	Supported
H_6	*Organizational support directly influences the effectiveness of performance measurement.*	Supported
H_7:	*There is a positive correlation between organizational support and external support.*	Supported

As explained above and seen in the table, six of seven study variables were supported according to the data at hand. Only the hypothesis 4, which is about the direct relationship of external support with effectiveness of performance measurement, was not supported. In the next section, the findings of the study, including the hypotheses testing results, and their implications will be discussed in detail.

CHAPTER 6: DISCUSSIONS, IMPLICATIONS, AND LIMITATIONS

In this chapter, the findings explained in the previous chapter will be discussed in detail, and their possible theoretical, methodological, managerial, and policy implications will be evaluated. Later in the chapter, limitations and possible future research directions will be provided.

6.1 Discussions

In this part of the chapter, the findings of the statistical analyses will be discussed. In the discussion, first the findings related to each latent variable of the study will be given. Then, the relationships among these variables, which are shown in the covariance structure model, will be discussed.

6.1.1 Organizational Support

Organizational support is a frequently mentioned factor which has an important influence in the quality of performance measures and performance measurement systems (Berman & Wang, 2000; Melkers & Willoughby, 2005; Sanger, 2008; Poister & Streib, 1999; Streib & Poister, 1999; Yang & Hsieh, 2007). It is an exogenous variable in the model and concerns the level of support from organizational actors, such as the mayor, top managers, lower level managers, and employees, for the use of performance measurement in the municipality. In the study, seven indicators, which are taken from the literature (Berman & Wang, 2000; de Lancer Julnes and Holzer, 2001), were used to measure the level of organizational support. In addition to direct questions regarding the support of mayor, top-managers, lower level managers, and employees, the study also used some indirect questions. These questions are regarding the frequency of the special meetings held in which performance measurement is discussed, the frequency of mayor's communication about the

importance of performance measurement to organizational actors, and the frequency of top-down internal (from managers to employees) communication about the issues related to performance measurement.

Since none of the indicators of organizational support had high correlation among each other, all of the indicators in the proposed model were kept in the generic measurement model. The confirmatory factor analysis showed that all of the indicators had statistical significance. Moreover, all of the factor loadings exceeded the predetermined criteria of .30. Therefore, none of the indicators were removed from the model. Goodness of fit statistics indicated that the measurement model of organizational support fits the data very well. According to the results of reliability analysis, the scale of organizational support has excellent reliability, with a Cronbach's alpha value of .90, meaning that it can produce consistent results at different times.

The results of the confirmatory factor analysis showed that, among the indicators of organizational support, *support of top-managers* (OS2) is the most important indicator with a standardized regression weight of .925, followed by *support of lower level managers* with a regression weight of .839, and *support of mayor* with a regression weight of .815. The results indicating that support of mayor is less important than support of managers may seem surprising at the first place, considering that Turkish municipalities are governed by a strong mayor government system. However, it should not be forgotten that even though mayor, along with the municipal council, is the one who defines political and administrative priorities, these are implemented by the close oversight and supervision of managers. Mayors are oftentimes are not involved in many aspects of the management. In that sense, managers have always the possibility to include their personal perceptions in the management of municipal services. A similar result can be inferred also from the comparison of indicators

*managers' communication about performance measuremen*t and *mayor's communication about performance measurement*. The former indicator has a regression weight of .625, whereas the latter has a regression weight of .555, indicating that managerial support is more indicative of organizational support than the support of mayor.

According to the descriptive analysis, support from employees for performance measurement was 63.6%, while support from mayors, top and lower level managers was ranging between 83.6% and 78.3%. Moreover, *support of employees* had a standardized regression weight of .581 in the confirmatory factor analysis which is considerably lower than those of support of other organizational actors. These results confirm the commonly accepted fact that the Turkish public administration works still in a highly top-down management system, in which the perceptions of employees play only a limited role in deciding about the important aspects of management.

Another interesting result of the descriptive analysis is that the questions asking directly the support of mayor and managers received approximately 80% positive statements (agree or strongly agree). On the other hand, indirect questions showed a lower level of support of the respondents. Regarding the communications of mayor and managers, 62.4% of the respondents reported that their mayor frequently emphasizes the importance of performance measurement, whereas only 55,9% reported such a frequent emphasis by the managers. The results show a discrepancy between the direct questions related to the support of the organizational actors and the indirect questions which aim to find out the extent these actors reflect their support for performance measurement on their practices.

6.1.2 External Support

External support, which is also an exogenous variable in the model, concerns the level of support from external actors, such as council members and citizens, for the use of performance measurement in the municipality. In the survey, respondents were asked five questions to understand the level of external support in their municipality. Similar to organizational support, the survey used indirect questions to understand the level of external support, in addition to direct questions regarding the support of these stakeholders. These indirect questions are regarding the extent the council members view performance measurement as an important aspect of decision making, the frequency of the council meetings in which performance measurement or data is discussed, and the extent citizens show their interest to municipality's performance information.

Since none of the indicators of external support had high correlation among each other, all of the indicators in the proposed model were kept in the generic measurement model. However, the model fit, even by correlating the measurement errors, could not be achieved. Therefore, the least important indicator for external support, *citizen interest in performance data* (ES5) was removed from the model. The confirmatory factor analysis showed that all of the indicators in the revised model had statistical significance and exceeded the predetermined criteria of .30. Goodness of fit statistics indicated that the revised measurement model of external support fits the data very well. According to the results of reliability analysis, the scale of external support has almost excellent reliability, with a Cronbach's alpha value of .88, meaning that it can produce consistent results at different times.

The results of confirmatory factor analysis show that all of the indicators have standardized regression weights ranging from .652 to .948, meaning that they all are

important factors influencing the external support of the municipality. According to the findings, *perceived importance by council members* (ES2) is the most important indicator, whereas *support of citizens* (ES4) is the least important one. That all of the indicators related to council members have higher standardized regression weights than those of the citizens related indicators shows that council members have a more significant role in shaping the activities of the municipalities, which is not surprising considering their position to take important decisions about the municipalities.

Despite the latest efforts in Turkey trying to achieve higher citizen participation in public administration, the existence of problems regarding participation is a well known fact accepted by researchers (Guney & Celenk, 2010; Guven, 2012; Kapucu & Palabiyik, 2008) and practitioners in Turkey. In that sense, this finding of the study is not surprising and confirmed the fact that citizens have still limited role in the design of public policies.

6.1.3 Technical Capacity

Technical capacity is an endogenous mediating variable in the study model. It explains the extent the organization can implement the performance measurement systems. In the literature, technical capacity of any organization is considered as one of the important factors of implementing successful performance measurement (Ammons & Rivenbark, 2008; Berman & Wang, 2000; de Lancer Julnes & Holzer, 2001; Streib & Poister, 1999; Wang & Berman, 2001). This capacity includes both the human resources capability and the technological capacity of the organization. In the study, technical capacity has six indicators, which are mostly adapted from Berman and Wang (2000) and Yang and Hsieh (2007). These indicators are the adequacy of the staff and information technology, the frequency the staff attending to training activities, and the competencies of the staff for the development of

quality performance measures, timely collection of performance data, and the analysis of performance data.

However, correlation analysis revealed that there is high correlation (.890) between two indicators, namely *competency in timely collection of performance data* (TC4) and *competency in performance data analysis* (TC5), which signs the high possibility of multicollinearity. As a result, TC5 was excluded from the model, since it had a lower regression weight.

The confirmatory factor analysis showed that all of the indicators in the revised model had statistical significance and exceeded the predetermined criteria of .30. Goodness of fit statistics indicated that the measurement model of technical capacity fits the data very well. According to the results of reliability analysis, the scale of technical capacity has excellent reliability, with a Cronbach's alpha value of .92.

The results of confirmatory factor analysis show that all of the indicators have standardized regression weights ranging from .724 to .936, meaning that they all are very important factors influencing the technical capacity of the municipality. According to the results, competencies of staff related to the implementation of performance measurement seem to be most important indicators of technical capacity. *Competency in performance measure development* (TC3) is the most important indicator with a standardized regression weight of .936 and *competency in timely collection of performance data* (TC4) is the following indicator with a regression weight of .904. The result, in a sense, shows that municipalities can more easily buy the physical material, find ordinary staff to fill the positions, and send these staff to training, however, it is very difficult to have or to create qualified staff for performance measurement activities. Qualified staff is the most discriminant factor of having or not having technical capacity. Since Turkish municipalities

are implementing performance measurement only for less than a decade, it may be difficult for them to find staff that is experienced and competent in this area. That having adequate information technology for the performance measurement being the least important indicator also confirms this logic in the sense that it is easier to acquire physical material than qualified personnel.

6.1.4 Quality of Performance Measures

Quality of performance measures is another endogenous mediating variable in the study. It is related to the adoption stage of performance measurement, and concerns the quality of performance measures. As mentioned earlier, this study uses Bouckaert's (1993) model in assessing the quality of performance measures. According to this model, performance measures can be considered as effective, if they are valid, legitimate, and functional. Putting it in SEM language, quality of performance measures is a second-order variable, which has three dimensions; validity, legitimacy, and functionality. As a result, measurement model for quality of performance measures includes these three dimensions.

In the study, first, measurement models of these three first-order variables were validated. Then, these models were combined in one model to form the measurement model of quality of performance measures. The indicators in the study regarding these dimensions were adapted from Streib and Poister (1999).

6.1.4.1 Validity. Validity of performance measures refers to the technical soundness of the measures. In the survey, respondents were asked six questions about the level of validity of performance measures developed and used in their municipality. These questions were about the extent that measures are developed from organizational missions, goals, and service standards, the extent they focus on what is important to measure (not the availability

of data), the extent they are up to date and clear, and the extent they track performance over time. However, correlation analysis revealed that there is high correlation (.906) between two indicators, namely between *derived from missions/goals* (V1) and *derived from service standards* (V2). Therefore, V2 was excluded from the model.

The results of CFA showed that all of the indicators in the revised model have statistical significance. Moreover, regression weights of all of the factor loadings exceeded the predetermined criteria of .30. Goodness of fit statistics indicated that the measurement model of validity fits the data very well. According to the results of reliability analysis, the scale of validity has excellent reliability, with a Cronbach's alpha value of .95.

CFA also showed that all of the indicators of *validity* have standardized regression weights ranging from .833 to .934, meaning that they all are very important factors influencing the validity of performance measures. According to the results, the most important indicator is *being clear/understandable* (V5) with a standardized regression weight of .934, followed by *measuring performance over time* (V6) with a regression weight of .901, whereas *focus on importance* (V3) is the least important indicator with a value of .833.

Results of the descriptive analysis showed that the positive statements regarding the validity of performance measures were mostly agreed or strongly agreed by more than the half of all respondents. Only the support for performance measures being up to date seems to be lower than the previous responses, yet there is more agreement (43.9%) than disagreement (34.8%) even on this statement. In sum, the responses for the validity of performance measures reveal a support from the respondents about the validity of their performance measures.

6.1.4.2 Legitimacy. Legitimacy is the second dimension of the latent variable, quality of performance measures. Legitimacy of performance measures is about the positive perceptions of stakeholders regarding the performance measures. The proposed measurement model for legitimacy has six indicators. For legitimacy, the respondents were asked the extent their managers, elected officials, and employees involve in the development of performance measures, and the extent these groups perceive performance measures which are developed by their municipality as useful.

According to correlation analysis, none of the indicators had high correlation among each other. Therefore, none of the indicators in the proposed model were removed from the model. The confirmatory factor analysis showed that all of the indicators in the revised model had statistical significance and exceeded the predetermined criteria of .30. Goodness of fit statistics indicated that the measurement model of legitimacy fits the data very well. According to the results of reliability analysis, the scale of legitimacy has excellent reliability, with a Cronbach's alpha value of .92.

The results of confirmatory factor analysis show that all of the indicators have standardized regression weights ranging from .760 to .888, meaning that they all are very important factors influencing the legitimacy of performance measures. According to the findings, the most important indicator is *perceived usefulness by elected officials* (L4) with a standardized regression weight of .888, followed by *perceived usefulness by managers* (L5) with a regression weight of .836, whereas *perceived usefulness by employees* has only a regression weight of .786. Considering the top-down management style of Turkish public administration, the results are not surprising. What employees perceive about something has only a limited role in affecting how activities will be conducted. A similar result can also be inferred from the least important indicator for legitimacy, which is *involvement of employees*

(L2) with a regression weight of .760. The results of the descriptive analysis also confirm that employees are less involved in the development of performance measures than managers are. A similar result was found by Streib and Poister (1999) who evaluated the design of performance measurement systems of the US municipalities with populations more than 25,000. They found that involvement of lower level employees and citizens in the design of performance measures is a very rare event. In general, the results of the confirmatory factors analysis seem to be in accordance with the criticism that top-down management system of Turkish public administration leaves very little room for employees to contribute to the improvement of public services (Kapucu & Palabiyik, 2008).

6.1.4.3 Functionality. The last dimension of the quality of a performance measure is its functionality. Functionality, in this sense, refers to the benefit creating potential of the performance measures. In the survey, respondents were asked five questions about the legitimacy of performance measures developed and used in their municipality. These questions were about the extent that performance measures developed in the municipality have the potential to improve service quality, decision-making capacity, employee motivation, organizational learning, and communication of managers with elected officials.

According to correlation analysis, none of the indicators had high correlation among each other. Therefore, none of the indicators in the proposed model were removed from the model. The confirmatory factor analysis showed that all of the indicators in the revised model had statistical significance and exceeded the predetermined criteria of .30. Goodness of fit statistics indicated that the measurement model of functionality fits the data very well. According to the results of reliability analysis, the scale of functionality has excellent reliability, with a Cronbach's alpha value of .96.

The results of confirmatory factor analysis show that all of the indicators have standardized regression weights ranging from .895 to .940, meaning that they all are very important factors influencing the functionality of performance measures. According to the table, the most important indicator is *potential for increasing employee motivation* (F3) with a standardized regression weight of .940, followed by *potential for stimulating organizational learning* (F4) with .919 and *potential for decision-making capacity improvement* (F2) with .910. The least important two indicators are *potential for service quality improvement* (F1) with a value of .895 and *potential for improving external communication* (F5) with .897. The results show that the benefits of performance measurement which benefit to the internal functions of the organization are more important for Turkish municipalities. This result is in accordance with the general acceptance that Turkish public administration still works with government perspective rather than a collaborative governance perspective. In general, government perspective focuses more on organizations' internal affairs and structure, whereas governance focuses on relationships with others, such as interorganizational, intergovernmental and crosssectoral relationships, which is mainly an external issue.

6.1.4.4 Integrated model. After validating the measurement models of three dimensions of quality of performance measures, next step was to integrate these models into one measurement model for quality of performance measures. As mentioned above, quality of performance measures is a second-order variable, which has three first-order variables; validity, legitimacy, and functionality. The integrated measurement model consists of three first-order variables and a total of 16 indicators.

Examination of the results shows that all of the indicators of quality of performance measures have critical ratios higher than 1.96 and p values lower than .05, which means that factor loadings for all indicators show statistical significance. The indicators have

standardized regression weights ranging from .738 to .934, meaning that they all are important factors influencing their respective first-order variables. Since all of the goodness of fit statistics, but model chi-square, showed very good fit for the model, the measurement model did not need any revision.

Among the three dimensions of quality of performance measures, legitimacy appeared to be the most indicative of quality with a standardized regression weight of .944, followed by validity with a weight of .910. According to the results, functionality is the least important dimension with a weight of .758. Examination of the descriptive analysis show that the responses regarding the indicators of functionality did not show high variation as the indicators of legitimacy and validity did. As a result, functionality dimension appeared to be less indicative of the quality of performance measures in the sample.

6.1.5 Effectiveness of Performance Measurement

Effectiveness of performance measurement is the endogenous variable of the study. It is related to the implementation phase of performance measurement. It is mainly about the effects of performance measurement and concerns the extent that performance measurement achieves its intended results and objectives. In the study, effectiveness of performance measurement was measured with ten indicators, which are mainly adapted from Yang and Hsieh (2007). The indicators were about to the extent performance measurement improves productivity and service quality, increases employee motivation, stimulates organizational learning, improves relations with the community, helps managers to identify the problems and create solutions for these problems, facilitates better decisions, and contributes to the managerial communication with elected officials.

However, correlation analysis revealed that there is high correlation between several indicators. In order to eliminate the high risk of multicollinearity, two indicators, namely *improvement in service quality* (E2) and *better solution of problems* (E8) were excluded from the model. As a result, generic model for effectiveness of performance measurement consisted of eight indicators.

Because of the problems with the goodness of fit statistics, the model was revised by correlating the measurement errors of the indicators. The results of the confirmatory factor analysis showed that all of the factor loadings of the indicators in the revised model are statistically significant and exceeded .30, showing that they are important factors for the effectiveness of performance measurement. Goodness of fit statistics indicated that the measurement model of effectiveness of performance measurement fits the data very well. According to the results of reliability analysis, the scale of effectiveness of performance measurement has excellent reliability, with a Cronbach's alpha value of .97.

The results of confirmatory factor analysis show that all of the indicators have standardized regression weights ranging from .817 to .942, meaning that they all are important factors influencing the effectiveness of performance measurement. According to the table, the most important indicator is *better decision-making* (E9) with a standardized regression weight of .942, followed by *stimulation of organizational learning* (E4) with .929 and *improvement in productivity* (E1) with .923.

Descriptive statistics regarding effectiveness of performance measurement also sheds light to one of the research questions of the study: "To what extent do Turkish municipalities implement performance measurement systems effectively?" The results show that most of the respondents, ranging from 76.1% to 64.2%, agreed or strongly agreed to these ten questions

indicating a clear support for the effectiveness of performance measurement in their municipalities.

Cost reduction (E6) is the least important indicator with a value of .817. Moreover, descriptive statistics also show that the statement about the cost reduction benefit of performance measurement received the least positive response rate (64.2%) among the indicators of quality of performance measures. The results show that the respondents have more hesitations about whether performance measurement practices of the municipality saves them some resources.

6.1.6 Covariance Structure Model

In this part, the SEM results regarding the covariance structure model will be discussed. SEM is an analytical method used to test the hypothesized relationships among the study variables. Validated measurement models explained in the previous parts were combined into one model, namely covariance structure model.

In the study, SEM is used to answer the research question of "whether or to what extent do quality of performance measures, technical capacity, organizational support, and external support have influence on the effectiveness level of performance measurement systems in Turkish municipalities?" In order to answer the research question, the hypothesized relationships between organizational support, external support, technical capacity, quality of performance measures and effectiveness of performance measurement were examined in a covariance structure model. Context-Design-Performance (CDP) model, which is heavily influenced from contingency theory and Donabedian's Structure-Process-Outcome model, was used as the main theoretical guidance in the study to examine the predictors of effective performance measurement in Turkish municipalities.

In the study, there are seven hypotheses about the relationships of the study variables. The first hypothesis is about the positive relationship between quality of performance measures and effectiveness of performance measurement. While the former is about the performance in the adoption stage, the latter is about the performance in implementation stage. The results of the SEM analysis show that quality of performance measures is positively and significantly related with effectiveness of performance measurement (β=.787, $p < 0.01$) as hypothesized in the study. This result, especially the high standardized regression weight, shows that in order for the effectiveness of performance measurement having high quality performance measures is a prerequisite. Having such a high relationship among these variables emphasizes the importance of technical and rational aspects on the effectiveness of performance measurement systems in the municipalities. Without creating and adopting appropriate measures it is not possible to accrue most of benefits of having performance measurement systems. The results are consistent with the literature, which similarly found a positive association between the level of adoption and level of implementation (de Lancer Julnes & Holzer, 2001; Yang & Hsieh, 2007).

The second hypothesis of the study is about the indirect relationship of technical capacity and effectiveness of performance measurement mediated by quality of performance measures. The results show that both the relationship between technical capacity and quality of performance measures (β=.736, $p < 0.01$), and the relationship between quality of performance measures and effectiveness of performance measurement (β=.787, $p < 0.01$) are positive and statistically significant, indicating that the hypothesis is supported. Multiplying the standardized regression weights of the paths give the indirect effect of technical capacity on effectiveness of performance measurement, which is .579. Having such a high relationship shows that technical capacity of any organization is one of the important factors of

implementing successful performance measurement, which is consistent with the literature (Ammons & Rivenbark, 2008; Berman & Wang, 2000; de Lancer Julnes & Holzer, 2001; Streib & Poister, 1999; Wang & Berman, 2001). Berman and Wang (2000) found that technical capacity is strongly associated with the increased use of performance measurement. In a follow up study, they also found a positive association between the professional competency and the use of both output and outcome measures (Wang & Berman, 2001). Similarly, de Julnes and Holzer (2001) found that resources, which they define similar to technical capacity, are positively and significantly related to the level of adoption of performance measures.

The third hypothesis of the study is about the indirect relationship between external support and effectiveness of performance measurement mediated by technical capacity and quality of performance measures. The results show that, in addition to the two relationships mentioned above, external support has a positive and statistically significant relationship with technical capacity (β=.388, $p < 0.01$), indicating that the hypothesis is supported. Indirect relationship between external support and effectiveness of performance measurement has a standardized regression weight of .225 (.388*.736*.787). The results confirm that external support has the potential to contribute to an increase in the technical capacity for performance measurement, since resources related to the capacity development in this area is highly dependent on the approval of elected officials (de Lancer Julnes & Holzer, 2001). Moreover, as argued by Yang and Wu (2013), citizen support both creates a pressure on the development of capacity and also increases the legitimacy of decisions taken in this regard.

The fourth hypothesis of the study is about the direct relationship between external support and effectiveness of performance measurement. Among the study hypotheses, this is the only hypothesis which was not supported by the data at hand. The results show that the

direct relationship of external support with effectiveness of performance measurement is not statistically significant, since the p value is .575, which is higher than the predetermined alpha level of .05. Although the direction of the relationship is positive as hypothesized, the effect is very trivial, and therefore, the hypothesis was not supported. This result is not consistent with the literature. Several studies showed that most of the important decisions, especially the ones with budgetary results, are taken with political incentives rather than objective/rational criteria (de Lancer Julnes & Holzer, 2001). Especially if the support of council members for performance measurement is low, it is always possible that council members may neglect the results of performance measurement and insist on taking decisions which may reduce the effectiveness of performance measurement systems.

Several reasons can account for such a result for the context of Turkish municipalities. First, citizen engagement in the activities of Turkish municipalities is lower compared to their counterparts in the US and the Europe. Citizens show their interest in the municipal activities only to the extent that their individual or communal well-beings are affected. They decide whether to re-elect the incumbent mayor based on their perceptions regarding the delivery of municipal services, but they have very little, if not any, power to affect how these services are delivered and the municipality is governed. Therefore, it is not very surprising that their level of support for performance measurement does not affect directly the effectiveness of performance measurement in the municipalities.

Secondly, council members are not as powerful in giving directions to the Turkish municipal organizations as their counterparts are. Turkish municipalities are governed by a strong mayor system. In this system, mayor is at the centre of all decision-making and implementation processes. In most cases, the majority of council members and the mayor are the members of the same political party. It is very rare that this majority opposes to the

general guidance of the mayor in the main decisions of the municipality. In that sense, the mayor's support for performance measurement is more important than that of council members. If mayor sticks to the requirements of performance measurement, low support for performance measurement among the council members might have a limited negative impact on the effectiveness of performance measurement.

The fifth hypothesis of the study is about the indirect relationship of organizational support and effectiveness of performance measurement mediated by technical capacity and quality of performance measures. The results show that, in addition to the two relationships discussed in the first two hypotheses, organizational support has a positive and statistically significant relationship with technical capacity (β=.242, $p < 0.01$), indicating that the hypothesis is supported. Indirect relationship between organizational support and effectiveness of performance measurement has a standardized regression weight of .140 (.388*.736*.787). The results are consistent with the literature stating that support from mayor, top management, lower level management, and employees contribute substantially to the successful implementation of performance measurement systems (Berman & Wang, 2000; Melkers & Willoughby, 2005; Sanger, 2008; Poister & Streib, 1999; Streib & Poister, 1999; Yang & Hsieh, 2007). Organizational support for performance measurement can contribute to the development of high quality performance measures by providing required resources for the development of technical capacity and also by motivating the personnel about the importance of performance measurement. In her study, Taylor (2006) found that measurement systems that receive the support of both higher and lower level employees are more likely to be better designed, implemented, and provide identifiable benefits for the organization (Taylor, 2006).

The sixth hypothesis is about the direct relationship between organizational support and effectiveness of performance measurement. The results of the SEM analysis indicate a positive and statistically significant direct relationship between these variables (β=.117, $p < 0.01$). Although the relationship is not strong, it is still significant. As a result, the hypothesis was supported. The results are consistent with the literature. Especially low levels of organizational support may create challenges for the effective implementation of performance measurement (Moynihan, 2005). Most of the literature does not differentiate between the direct and indirect effects of organizational support on the effectiveness of performance measurement. But, they are mostly in accordance that it is very important. Yang and Hsieh (2007) found in their study that organizational support is the most important predictor of effectiveness in both the adoption and implementation stages of performance measurement. In a similar vein, Folz et al. (2009) found that the lack of organizational support is the single most important factor that can explain why performance measurement fell short of meeting the expectations of chief executives.

The seventh, and the last hypothesis, is about the correlation between two variables of support. The results also indicate that there is a positive and significant correlation among organizational support and external support with a standardized regression weight of .596. The results are consistent with the literature. Yang and Hsieh (2007) also found a similar relationship among these variables.

Among the two control variables of the study, only population had a statistically significant relationship with effectiveness of performance measurement (β=-.120, $p < 0.01$). However, the direction of the relationship between population and effectiveness of performance measurement is negative, meaning that municipalities with higher populations tend to have less effective performance measurement practices, which is surprising. It might

be expected that larger cities with more resources and technical capacity would do better in the effectiveness of performance measurement. The other control variable, type of the municipality, did not show a significant relationship with effectiveness of performance measurement (β=.050, p =.106), which can be understandable, since, in Turkey, both human and financial resources of municipality are positively related to the population of the municipalities rather than the type of the municipality. For example, a metropolitan district municipality with a population of 1,000,000 has more resources than a city municipality which has a population of 50,000. Having different levels of resources may be more related to effectiveness of performance measurement than the type of municipality.

6.2 Implications

6.2.1 Theoretical Implications

In this study, context-design-performance (CDP) model, which is heavily influenced from contingency theory (Agiro, 2011; Burke & Litwin, 1992; Goltz, 2006; Marathe, 2006; Wan 1995) and Donabedian's structure, process, and outcome model (Agiro, 2011; Goltz, 2006; Wan, 2002), was used as the main theoretical guidance to examine the predictors of effective performance measurement in Turkish municipalities. In this model, contextual factors influence organizational performance both directly and indirectly (via design factors). Other than the contextual factors, design factors also influence organizational performance (Wan, 1995). This model does not focus only on an organization's internal operations and it is especially applicable when external factors play an important role in the performance of organization (Wang, 2010).

In the literature, external and organizational support for the use of performance measurement and technical capacity of organizations for performance measurement have

been noted as some of the most important factors which influence the implementation of performance measurement (Berman & Wang, 2000; Broad, 2006; de Lancer Julnes & Holzer, 2001; Fernandez & Rainey, 2006; Hatry, 2006; Poister & Streib, 1999; Streib & Poister, 1999; Taylor, 2006; Wang & Berman, 2001; Yang & Hsieh, 2007). Therefore, these variables are included in the model as the predictors of effective performance measurement. In this study, external support and organizational support have been considered as the contextual factors, and technical capacity as the design factor. Having external support as the contextual factors conforms to the propositions of contingency theory, which posit that external factors have influence in the organizational design, structure, or performance. Inclusion of organizational support and technical capacity in the model as contextual and design factors respectively is in accordance with the main idea of Donabedian's SPO model that better structure and process leads to a better performance.

Regarding the performance part of the CDP model, the study used the construct of effectiveness, which shows the performance of an organizational practice, which is performance measurement in this case. In other words, the study concerned the performance of performance measurement. Based on the explanations of de Lancer Julnes and Holzer (2001), which see the utilization of performance measurement systems consisting of two stages, namely adoption of performance measures and implementation of performance measurement, the study included quality and effectiveness in these two stages as the variables of performance in the model.

In general, the results of the study showed consistency with the propositions of theoretical models and theories used in the study. The study confirmed that the organizational contextual factor of the study, organizational support, has both a direct and indirect relationship with the quality of performance measures, which is in accordance with the CDP

and SPO models. On the other hand, the environmental contextual factor of the study, external support, has only a direct relationship, but not an indirect one. The result is consistent with contingency theory, which emphasizes the importance of fit between external and organizational factors in influencing the performance of organization. However, according to the CDP model both direct and indirect relationship between the context factor and performance could be expected. That the design factor of the model, technical capacity, has a positive and significant relationship with the performance variables of the study is also in accordance with CDP and SPO models.

6.2.2 Methodological Implications

One of the important methodological implications of the study is related to the operationalization of the study variables. None of the scales of latent constructs were taken directly from a previous source, but they were compiled and adapted from several sources. The results of confirmatory factor analysis and reliability analysis show that the scales have high levels of validity and reliability. Only concern may be that the questions regarding the each construct were tried to be minimized as possible, since the inflation in the number of the questions could deter the response rate of the study and increase the missing values. In that sense, especially the scale of external support should be revisited for the future studies. The scales of organizational support, technical support, and effectiveness of performance measurement can be used in the other studies without making any changes. The scale of effectiveness for performance measures can also be used without a change in the studies with complex models. However, if the study model is not complex, this scale can be also revisited and the numbers of the indicators may be increased.

The utilization of a scale for quality of performance measures is also another methodological strength of the study. In order to examine the adoption level of performance

measurement, previous studies (Berman & Wang, 2000; de Lancer Julnes & Holzer, 2001; Yang & Hsieh, 2007) used only the depth of performance measurement by trying to find out the extent that some kind of performance measures, such as output, outcome, and quality, are used in various functions of the organization. Although this is an acceptable method, using Bouckaert's three dimensional quality of performance measures construct has the advantage of drawing a more complete picture of the reality regarding the quality of the performance measures used in the organization.

Another methodological implication of the study is the use of the online survey tool, Qualtrics, for collecting the data from a wide range of samples which are physically dispersed around the country. Using this tool saves considerable time and resources for the researcher. Moreover, the results of the surveys can easily be transformed to a SPSS file and become ready for the data analysis. Furthermore, the utilization of the online survey tool made it easier to send the survey to only those who did not respond to the survey in the previous waves. By using the advantages of this online tool, the survey was sent to the respondents four times, which increased the response rate of the study.

The survey was constructed by using the literature in English, but implemented in Turkish. In order to reduce the problems that emerge from cultural differences the translation of the questions were made based on their functional meanings, not the literal ones. Moreover, the translation was controlled by several Turkish practitioners and academicians, who are familiar with both the literature in the US and the public administration in Turkey. Revisions are made according to their recommendations. As a result, it is important to pay attention to the cultural differences when constructing a survey using the literature of a different culture.

6.2.3 Managerial and Policy Implications

The results of the SEM analysis show that quality of performance measures is the variable which has the highest impact (β=.787) on effectiveness of performance measurement, which is consistent with the findings in the literature (de Lancer Julnes & Holzer, 2001; Yang & Hsieh, 2007). This result shows that in order for the effectiveness of performance measurement having high quality performance measures is a prerequisite, which is only possible with high levels of technical capacity.

The results of the descriptive analysis regarding technical capacity show that while most of the respondents agreed that their municipalities have adequate information technology, they reported significant problems regarding the adequacy of the staff, some of their competencies related to performance measurement, and the frequency their staff attend to the training activities. Considering the high effect (β=.579) of technical capacity on effectiveness of performance measurement, which is consistent with the literature (Ammons & Rivenbark, 2008; Berman & Wang, 2000; de Lancer Julnes & Holzer, 2001; Streib & Poister, 1999; Wang & Berman, 2001), it is important for the managers and elected officials to invest more on technical capacities of the municipalities.

The results of the descriptive analysis confirm that employees are less involved in the development of performance measures than managers are. Only 39% of the respondents agreed or strongly agreed that employees are involved in the development process. Having acknowledged the benefits of employee support, Holzer and Yang (2004) also mention the importance of employee participation in the design and improvement of performance measurement systems for enhancing the employee buy-in of these systems. As argued by Häggroth (2013), key to success in the introduction of new practices is the inclusion of people working in the organization to the each step of the change process. Since performance

measurement is one of such practices for Turkish municipalities, it is important that employees are convinced to take active part in the design of it. However, despite the changes in recent years, Turkish administrative culture can still be considered as lacking participatory perspectives. Some managers and especially most of lower level employees seem not to be involving in the development process of measures, which may negatively influence their perceptions about usefulness of those performance measures. Consequently, the results of the study create additional support for the argument that the managers should put more effort in the involvement of employees in the development of performance measures

The results of the SEM analysis also confirmed that both organizational support and external support, despite the latter has only an indirect relationship, are important predictors of effectiveness of performance measurement as hypothesized in the study. However, close examination of the descriptive analysis reveals that the support of employees and citizens are not high as the other stakeholders. According to the descriptive analysis, support from employees for performance measurement was 63.6%, while support from mayors, top and lower level managers was ranging between 83.6% and 78.3%. Moreover, support of employees had a standardized regression weight of .581 in the confirmatory factor analysis which is considerably lower than those of support of other organizational actors. These results confirm the commonly accepted fact that the Turkish public administration works still in a highly top-down management system, in which the perceptions of employees play only a limited role in deciding about the important aspects of management. However, it is a well known and generally acknowledged fact that the resistance of employees against change may create significant challenges for top managers and sabotage the success of the performance systems (Berman & Wang, 2000; Poister & Streib, 1999; Streib & Poister, 1999; Taylor,

2006). Therefore, it is important that new strategies to be developed to increase the support of employees for performance measurement systems utilized in Turkish municipalities.

Despite the latest efforts in Turkey trying to achieve higher citizen participation in public administration, the existence of problems regarding participation is a well known fact accepted by researchers (Guney & Celenk, 2010; Guven, 2012; Kapucu & Palabiyik, 2008) and practitioners in Turkey. In that sense, this finding of the study regarding the relatively low levels of citizen support for performance measurement is not surprising and it confirmed the fact that citizens have still limited role in the design of public policies. However, as argued by Yang and Wu (2013), citizen support both creates a pressure on the development of capacity and also increases the legitimacy of decisions taken in this regard. An important way of increasing the citizen support for the municipal practices is to encourage their participation in the design of these practices. According to Nalbandian, O'Neill, Wilkes, and Kaufman (2013), citizen participation is not optional any more, but imperative as being an important way of bridging the gap between political acceptability and administrative sustainability, which increases the legitimacy of governing institutions. However, engaging citizens with traditional local government practices is one of the three leadership challenges local governments face today (Nalbandian et al., 2013). Therefore, it is important that the policies regarding the citizen participation should be revised and an increased participation is promoted in public administration including the design of performance measurement systems of the municipalities.

6.3 Limitations

Despite the expected benefits of this study explained in the previous sections, it should not be ignored that this study has some limitations. Most importantly, the study gathers data by a self-administered survey. The fact that the suggested relations can be

examined only based on the perceptions of the respondents may create some validity problems. It is possible that the person who responded on behalf of the municipality may have limited knowledge about the municipality's performance measurement activities and their results, or may prefer to answer in a more favorable way for himself/herself instead of actually telling his real opinion. Moreover, it is possible that survey respondents may understand the questions and concepts in the survey differently.

Another limitation of the study is that it depends on cross-sectional data. As argued by Yang & Hsieh (2007), with cross-sectional data, SEM shows mostly the associations between the study variables, but cannot guarantee causality. In order to mitigate this problem, this study carefully identified the hypotheses of the study based on the guidance of the theoretical framework and the literature. Nevertheless, for the future studies, a longitudinal study design of the same topic may better confirm causality and contribute to a better explanation of the relationships between the study variables.

6.4 Future Studies

In the study, four variables, organizational support, external support, technical capacity and quality of performance measures, were examined as the predictors of the effective performance measurement. However, in the literature many other variables are mentioned as the predictors of this construct. Future studies may examine the effects of other predictors on the effectiveness of performance measurement.

Secondly, this study model is a unique model, which has not been used in any other study before. Although the analyses showed that the model has high validity, reliability and hypothesized relationships were mostly supported, it is important that this study is replicated

in other countries. Such a replication of the study model would contribute to the generalizability of the study.

As mentioned in the limitations section, the study gathered data by a self administered survey. The utilization of the survey contributed to the generalizability of the results at least in Turkey. However, it also created some validity problems. Instead of examining a wide sample of municipalities, the future studies may concentrate on a small group of municipalities and examine the real effects of performance measurement, not the perceived ones, by using other data collection techniques which mitigate the researchers' dependence on perceived data. In that sense, agency records can be examined and interviews may be conducted to better understand the opinions of the stakeholders.

This study found that support of employees and citizens for the performance measurement practices of the municipalities are relatively low compared to that of the other stakeholders. Future studies may explore the factors which can increase the levels of them. Such a study may contribute to an increase in the effectiveness of performance measurement in the municipalities, which may increase productivity, and service quality.

APPENDIX A: SURVEY QUESTIONNAIRE

Survey Questionnaire

Title of Project: Predictors of an Effective Performance Measurement System: Evidence from Municipalities in Turkey

Principal Investigator: Sedat Eliuz

Faculty Supervisor: Naim Kapucu, PhD

You are being invited to take part in a research study. Whether you take part is up to you.

This survey aims to delineate factors that are important for the effectiveness of performance measurement systems. This survey will be used to identify the level of stakeholder (external and organizational) support for performance measurement, technical capacity of municipalities regarding performance measurement, effectiveness of created performance measures, and effectiveness of performance measurement systems in Turkish municipalities. You will be asked to respond to the survey questions online. The survey takes about 10-15 minutes to complete. Your responses are confidential, and will not be revealed without your consent; only aggregate results will be made available.

You must be 18 years of age or older to take part in this research study.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints, contact to Sedat Eliuz, Graduate Student, College of Health and Public Affairs, +1 (407) 965-7148 by email at sedateliuz@icisleri.gov.tr or Dr. Naim Kapucu, Professor of Public Administration, College of Health and Public Affairs, +1 (407) 823-6096 or by email at kapucu@ucf.edu.

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at +1 (407) 823-2901.

By clicking on the survey link below and completing the survey, you are consenting to take part in this study.

1. Organizational Support

Please assess the following statements regarding the level of organizational support for the performance measurement in your organization. Please use the following scale:

Strongly disagree	*Disagree*	*Neither agree nor disagree*	*Agree*	*Strongly agree*
1	*2*	*3*	*4*	*5*

[ ] Our mayor supports the use of performance measurement in our municipality.

[ ] Department heads support the use of performance measurement in our municipality.

[ ] Lower-level (middle) managers support the use of performance measurement in our municipality.

[ ] Employees support the use of performance measurement in our municipality.

[ ] Special meetings regarding performance measurement are frequently organized in our municipality.

[ ] Our mayor frequently emphasizes the importance of performance measurement by showing his/her interest in performance measurement in his/her communication to organizational actors (managers and employees).

[ ] Our managers regularly emphasize the importance of performance measurement by showing their interest in performance measurement in their communication to employees.

2. External Support

Please assess the following statements regarding the level of external support for the performance measurement in your organization. Please use the following scale:

Strongly disagree	*Disagree*	*Neither agree nor disagree*	*Agree*	*Strongly agree*
1	*2*	*3*	*4*	*5*

[ ] Council members support the use of performance measurement in our municipality.

[ ] Council members view performance measurement as an important aspect of decision making (resource allocation, strategic planning, etc.)

[ ] Performance measurement or performance data are discussed frequently in council meetings of the municipality.

[ ] There is a considerable citizen support for the use of performance measurement in our municipality.

[ ] Citizens show their interest to our performance information.

3. Technical Capacity

Please assess the following statements regarding the level of technical capacity for the performance measurement in your organization. Please use the following scale:

Strongly disagree	*Disagree*	*Neither agree nor disagree*	*Agree*	*Strongly agree*
1	*2*	*3*	*4*	*5*

Our departments in our municipality;

[ ] have adequate number of staff for performance measurement.

[ ] have adequate information technology for performance measurement.

[ ] have staff capable of developing high quality performance measures.

[ ] have staff capable of collecting performance data in a timely manner.

[ ] have staff capable of analyzing performance data.

[ ] have staff attending regularly to conferences/workshops/trainings related to performance measurement.

4. Quality of performance measures

4.1. Validity of Performance Measures

Please assess the following statements regarding the validity of performance measures in your organization. Please use the following scale:

Strongly disagree	*Disagree*	*Neither agree nor disagree*	*Agree*	*Strongly agree*
1	*2*	*3*	*4*	*5*

[ ] Our performance measures are mostly derived from the mission, goals, and objectives established for our programs and/or departments.

[ ] Our performance measures are mostly derived from the service standards established for our programs and/or departments.

[ ] When developing performance measures, we focus on what is important to measure rather than the availability of data.

[ ] Our performance measures current and up to date.

[ ] Our performance measures are clear and not confusing for the organizational actors (managers and employees).

[ ] We use our performance measures to track performance of our municipality over time.

4.2. Legitimacy of Performance Measures

Please assess the following statements regarding the legitimacy of performance measures in your organization. Please use the following scale:

Strongly disagree	*Disagree*	*Neither agree nor disagree*	*Agree*	*Strongly agree*
1	*2*	*3*	*4*	*5*

[ ] Managers involve in the development of performance measures.

[ ] Lower level employees involve in the development of performance measures.

[ ] We keep the city council informed about our efforts to develop performance measures.

[ ] Elected officials consider that our performance measures are useful.

[ ] Managers consider that our performance measures are useful.

[ ] Lower level employees consider that our performance measures are useful.

4.3. Functionality of Performance Measures

Please assess the following statements regarding the functionality of performance measures in your organization. Please use the following scale:

Strongly disagree	*Disagree*	*Neither agree nor disagree*	*Agree*	*Strongly agree*
1	*2*	*3*	*4*	*5*

[ ] Our performance measures have the potential to improve service quality.

[ ] Our performance measures have the potential to improve the quality of decisions or decision-making capacity.

[ ] Our performance measures have the potential to improve the level of employee motivation.

[ ] Our performance measures have the potential to stimulate organizational learning.

[ ] Our performance measures have the potential to improve the communication between managers and elected officials.

5. Effectiveness of Performance Measurement

Please assess the following statements regarding the effectiveness of performance measurement in your organization. Please use the following scale:

Strongly disagree	*Disagree*	*Neither agree nor disagree*	*Agree*	*Strongly agree*
1	*2*	*3*	*4*	*5*

[ ] The use of performance measurement improves productivity in our municipality

[ ] The use of performance measurement improves service quality in our municipality.

[ ] The use of performance measurement increases employee motivation in our municipality.

[ ] The use of performance measurement stimulates organizational learning by giving feedbacks about the performance of departments in our municipality.

[ ] The use of performance measurement improves our relations with the community.

[ ] The use of performance measurement creates reductions in the costs of our municipal services.

[ ] The use of performance measurement helps our managers to better identify managerial and operational problems in our departments.

[ ] The use of performance measurement helps our managers to better develop solutions to managerial and operational problems in our departments.

[ ] Performance measurement can help managers make better decisions.

[ ] This organization's performance measurement helps managers communicate more effectively with elected officials.

6. Open Ended Questions:

Are there any other factors, which have not been covered in this survey, that you think are very important in influencing the effectiveness of performance measurement system in your municipality? Please specify.

What are the main obstacles that limit the effectiveness of performance measurement system in your municipality? Please explain.

Can you name of some of the main activities that your municipality carries out in order to increase the effectiveness of performance measurement system?

---

7. Control Variables

The population of your municipality?

a) lower than 10.000 b) between 10.001 and 50.000 c) between 50.001 and 100.000 d) between 100.001 and 250.000 e) higher than 250.000

Type of your municipality?

a) Metropolitan b) Metropolitan district c) Province d) District

8. About the Respondent

What is your position?

a) Mayor b) Department Head c) Middle Manager d) Specialist e) Other staff (Please specify)

How long have you been working in this municipality?

a) less than 1 year b) between 1 and 3 years c) between 3 and 10 years d) more than 10 years

How long have you been working in performance measurement activities of your municipality?

a) less than 1 year b) between 1 and 3 years c) between 3 and 10 years d) more than 10 years

Thank you very much for your participation!

APPENDIX B: SURVEY QUESTIONNAIRE IN TURKISH

ANKET

Projenin Adı: Etkili Bir Performans Ölçüm Sistemini Sağlayan Unsurlar: Türk Belediyeleri Üzerinde Bir İnceleme

Araştırmacı: Sedat Eliuz

Danışman: Naim Kapucu

Belediyelerdeki performans ölçüm sistemlerinin etkinliğini konu alan çalışmamıza hoş geldiniz. Bu çalışmaya katılım tamamen gönüllülük esasına göredir.

Bu anket yerel yönetimlerin performans ölçüm sistemlerinin etkililiği için önem arzeden faktörlerin belirlenmesine yardımcı olmayı amaçlamaktadır. Anket, Türk belediyelerinde performans ölçümüne kurumsal ve dış destek, performans ölçümü konusundaki belediyenin teknik kapasitesi, geliştirilen performans kriterlerinin etkililiği ve genel olarak performans ölçüm sistemlerinin etkililiğini ölçmek için kullanılacaktır. Anket sorularına online olarak cevap vermeniz beklenmektedir. Anketin yaklaşık olarak 10-15 dakika içinde tamamlanabileceği düşünülmektedir.

Bu çalışmaya katılmak için 18 yaşından büyük olmanız gerekmektedir.

Çalışma ile ilgili soru ve şikâyetler için iletişim bilgileri: Bu çalışmayla ilgili başka soru, endişe ya da şikâyetleriniz var ise, UCF College of Health and Public Affairs'de doktora yapmakta olan Sedat Eliuz'a (+1) 407-965-7148 numaralı telefondan veya sedateliuz@icisleri.gov.tr mail adresinden ulaşabilirsiniz. Ayrıca araştırmanın danışmanı Prof. Dr. Naim Kapucu (UCF Public Administration, College of Health and Public Affairs) ile (+1) 407-823-6096 numaralı telefon veya kapucu@ucf.edu mail adresinden irtibat kurabilirsiniz.

Çalışma ile ilgili hak ve şikâyetleriniz için IRB iletişim bilgileri: University of Central Florida'da insan katılımı ile yapılan araştırmalar Institutional Review Board (UCF IRB) gözetiminde yapılmaktadır. Bu çalışma IRB tarafından incelenmiş ve onaylanmıştır. Haklarıyla ilgili daha fazla bilgi sahibi olmak isteyenler, IRB ile Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 adresinden veya (+1) 407-823-2901 numaralı telefondan irtibat kurabilirler.

Aşağıdaki anket linkini tıklayarak ve anketi doldurarak bu çalışmada yer almayı kabul etmiş oluyorsunuz.

1. Kurumsal Destek

Lütfen aşağıdaki ifadeleri belediyenizin performans ölçümü konusundaki kurumsal desteği kapsamında değerlendirin. Lütfen aşağıdaki ölçeğe göre cevaplayın:

Tamamen katılmıyorum	*Katılmıyorum*	*Ne katılıyorum ne de katılmıyorum*	*Katılıyorum*	*Tamamen katılıyorum*
1	*2*	*3*	*4*	*5*

[ ] Belediye başkanımız belediyemizde performans ölçümünün kullanılmasını destekler.

[ ] Üst düzey yöneticilerimizin (genel sekreter, daire başkanı vb.) çoğu belediyemizde performans ölçümünün kullanılmasını destekler.

[ ] Orta düzey yöneticilerimizin (şube müdürleri, şefler vb.) çoğu belediyemizde performans ölçümünün kullanılmasını destekler.

[ ] Belediye çalışanlarının çoğu belediyemizde performans ölçümünün kullanılmasını destekler.

[ ] Belediyemizde yapılan toplantılarda performans ölçümü konusu sıklıkla gündeme gelir.

[ ] Belediye başkanımız performans ölçümüne olan ilgisini belediye görevlileriyle olan görüşmelerinde sıklıkla dile getirir.

[ ] Yöneticilerimizin çoğu çalışanlarla olan görüşmelerinde performans ölçümüne ilgisini düzenli olarak gösterir.

2. Kurum Dışı Destek

Lütfen aşağıdaki ifadeleri belediyenizin performans ölçümü konusundaki dış desteği kapsamında değerlendirin. Lütfen aşağıdaki ölçeğe göre cevaplayın:

Tamamen katılmıyorum	*Katılmıyorum*	*Ne katılıyorum ne de katılmıyorum*	*Katılıyorum*	*Tamamen katılıyorum*
1	*2*	*3*	*4*	*5*

[ ] Belediye Meclis üyelerinin çoğu belediyemizde performans ölçümünün kullanılmasını destekler.

[ ] Belediye Meclis üyeleri, kaynak tahsisi ve stratejik planlama gibi temel konularda kararlar alırken performans ölçümünü (bilgilerini) göz önünde bulundurulması gereken önemli bir husus olarak görürler.

[ ] Performans ölçümü ya da performans bilgileri belediye meclis toplantılarında sıklıkla görüşülür.

[ ] Belediyemizde performans ölçümünün kullanılması konusunda önemli ölçüde vatandaş desteği bulunmaktadır.

[ ] Vatandaşlar belediyemizin/birimlerimizin performans bilgilerine ilgilerini çeşitli yollarla (bilgi talebi, web sayfasını ziyaret vb.) göstermektedir.

3. Teknik Kapasite

Lütfen aşağıdaki ifadeleri belediyenizin performans ölçümü konusundaki belediyenizin teknik kapasitesi kapsamında değerlendirin. Lütfen aşağıdaki ölçeğe göre cevaplayın:

Tamamen katılmıyorum	*Katılmıyorum*	*Ne katılıyorum ne de katılmıyorum*	*Katılıyorum*	*Tamamen katılıyorum*
1	*2*	*3*	*4*	*5*

Belediyemizdeki dairelerin çoğunda;

[ ] performans ölçümü için yeterli sayıda personel bulunmaktadır.

[ ] performans ölçümü için yeterli iletişim teknolojisi imkânları bulunmaktadır.

[ ] etkili performans kriterleri geliştirebilecek yeterlilikte personel bulunmaktadır.

[ ] performans bilgilerini zamanı içinde toplayabilecek yeterlilikte personel bulunmaktadır.

[ ] performans bilgilerini analiz edebilecek yeterlilikte personel bulunmaktadır.

[ ] performans ölçümüyle ilgili konferanslara/çalıştaylara/eğitimlere düzenli olarak katılan personel bulunmaktadır.

4. Performans Kriterlerinin Kalitesi

4.1. Performans Kriterlerinin Doğruluğu

Lütfen aşağıdaki ifadeleri belediyenizin performans kriterlerinin teknik olarak doğruluğu kapsamında değerlendirin. Lütfen aşağıdaki ölçeğe göre cevaplayın:

Tamamen katılmıyorum	*Katılmıyorum*	*Ne katılıyorum ne de katılmıyorum*	*Katılıyorum*	*Tamamen katılıyorum*
1	*2*	*3*	*4*	*5*

[ ] Performans kriterlerimiz, doğrudan belediyemizin/dairelerimizin misyon, amaç ve hedefleri esas alınarak oluşturulmuştur.

[ ] Performans kriterlerimiz, belediyemizin/dairelerimiz için oluşturulmuş olan hizmet standartları esas alınarak oluşturulmuştur.

[ ] Performans kriteri oluştururken kolay bilgi toplayabileceğimiz alanlardan ziyade bizim için önemli olan alanlara yoğunlaşırız.

[ ] Performans kriterlerimiz sürekli ve düzenli olarak güncellenmektedir.

[ ] Performans kriterlerimiz açıkça anlaşılabilir olup kafa karıştırıcı değildir.

[ ] Performans kriterlerini dairelerin/birimlerin performanslarını zaman içinde takip etmek için kullanıyoruz.

4.2. Performans Kriterlerinin Meşruluğu

Lütfen aşağıdaki ifadeleri belediyenizin performans kriterlerinin meşruluğu kapsamında değerlendirin. Lütfen aşağıdaki ölçeğe göre cevaplayın:

Tamamen katılmıyorum	*Katılmıyorum*	*Ne katılıyorum ne de katılmıyorum*	*Katılıyorum*	*Tamamen katılıyorum*
1	*2*	*3*	*4*	*5*

[ ] Performans kriterlerinin oluşturulmasına yöneticilerin çoğu katılmaktadır.

[ ] Performans kriterlerinin oluşturulmasına yönetici olmayan çalışanların çoğu katılmaktadır.

[ ]Performans kriterlerini oluşturma çalışmalarımız konusunda belediye meclis üyelerini belli aralıklarla bilgilendirmekteyiz.

[ ] Belediyemizin seçimle gelmiş görevlilerinin çoğu belediyemizce oluşturulan performans kriterlerinin faydalı olduğunu düşünmektedir.

[ ] Yöneticilerin çoğu belediyemizce oluşturulan performans kriterlerinin faydalı olduğunu düşünmektedir.

[ ] Alt düzey çalışanlarımızın çoğu belediyemizce oluşturulan performans kriterlerinin faydalı olduğunu düşünmektedir.

4.3. Performans Kriterlerinin Fonksiyonel Olması

Lütfen aşağıdaki ifadeleri belediyenizin performans kriterlerinin fonksiyonelliği kapsamında değerlendirin. Lütfen aşağıdaki ölçeğe göre cevaplayın:

Tamamen katılmıyorum	*Katılmıyorum*	*Ne katılıyorum ne de katılmıyorum*	*Katılıyorum*	*Tamamen katılıyorum*
1	*2*	*3*	*4*	*5*

[ ] Belediyemizce oluşturulan performans kriterlerinin hizmet kalitesini artırma potansiyeli vardır.

[ ] Belediyemizce oluşturulan performans kriterlerinin belediye yöneticilerinin daha yerinde kararlar almasını sağlayabilir.

[ ] Belediyemizce oluşturulan performans kriterlerinin çalışanların motivasyonunu artırma potansiyeli vardır.

[ ] Belediyemizce oluşturulan performans kriterlerinin kurumsal öğrenmeyi teşvik etme potansiyeli vardır.

[ ] Belediyemizce oluşturulan performans kriterlerinin yöneticiler ve seçimle gelmiş görevliler arasındaki iletişimi iyileştirme potansiyeli vardır.

5. Performans Ölçümünün Etkililiği

Lütfen aşağıdaki ifadeleri belediyenizin performans ölçümünün etkililiği kapsamında değerlendirin. Lütfen aşağıdaki ölçeğe göre cevaplayın:

Tamamen katılmıyorum	*Katılmıyorum*	*Ne katılıyorum ne de katılmıyorum*	*Katılıyorum*	*Tamamen katılıyorum*
1	*2*	*3*	*4*	*5*

[ ] Performans ölçümü belediyemizce üretilen hizmetleri artırmaktadır.

[ ] Performans ölçümü belediyemizde hizmet kalitesini artırmaktadır.

[ ] Performans ölçümü belediyemizde çalışanların motivasyonunu artırmaktadır.

[ ] Performans ölçümü belediyemizde (dairelerin/birimlerin yüksek ve düşük performansları hakkında bilgi vererek) kurumsal öğrenmeyi teşvik etmektedir.

[ ] Performans ölçümünün kullanılması toplum ile olan ilişkilerimizi geliştirmektedir.

[ ] Performans ölçümünün kullanılması belediye hizmetlerinin maliyetlerinin azalmasına katkı sağlamaktadır.

[ ] Performans ölçümünün kullanılması yöneticilerimizin birimlerimizdeki yönetimle ve uygulamayla ilgili sorunları daha iyi tespit etmelerine katkı sağlar.

[ ] Performans ölçümünün kullanılması yöneticilerimizin birimlerimizdeki yönetimle ve uygulamayla ilgili sorunlara daha iyi çözümler bulmasına katkı sağlar.

[ ] Performans ölçümü yöneticilerimizin daha iyi kararlar alabilmesini sağlamaktadır.

[ ] Performans ölçümü yöneticilerimizin belediyemizdeki seçilmiş olan görevlilerle (belli konuları sayılarla ifade ederek)) daha etkili iletişim kurmalarına katkı sağlar.

6. Açık Uçlu Sorular

- Belediyenizdeki performans ölçme sisteminin etkililiğini artırdığını düşündüğünüz ama bu ankette yer almayan başka faktörler var mıdır? Lütfen açıklayınız.
  ……
- Belediyenizdeki performans ölçme sisteminin etkililiğini kısıtlayan ana engeller nelerdir?
  ……
- Belediyenizin, kullanmış olduğunuz performans ölçme sisteminin etkililiğini artırmak amacıyla yapmış olduğu temel faaliyetleri kısaca açıklayabilir misiniz?
  ……

6. Kontrol Değişkenleri

- Belediyenizin nüfusu ne kadardır?
  a) 10.000'den az b) 10.000 ve 50.000 arasında c) 50.001 ve 100.000 arasında
  d) 100.001 ve 250.000 arasında e) 250.000'den fazla
- Belediyeniz ne tür bir belediyedir?
  a) Büyükşehir b) Büyükşehir İlçe c) İl d) İlçe

7. Katılımcı Hakkında

- Belediyedeki göreviniz nedir?

  a) Belediye Başkanı b) Daire Başkanı c) Orta Düzey Yönetici (Şube Müdürü, Şef vb.) d) Uzman e) Diğer

- Bu Belediyede ne kadar süredir görev yapıyorsunuz?
  a) 1 yıldan az b) 1-3 yıl arası c) 3-10 yıl arası d) 10 yıldan fazla

- Belediyenizin performans ölçümü çalışmalarında ne kadar süredir görev yapıyorsunuz?
  a) 1 yıldan az b) 1-3 yıl arası c) 3-10 yıl arası d) 10 yıldan fazla

Katılımınız için teşekkür ederim!

APPENDIX C: THE IRB APPROVAL

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Exempt Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Sedat Eliuz**

Date: **June 16, 2014**

Dear Researcher:

On 6/16/2014, the IRB approved the following activity as human participant research that is exempt from regulation:

Type of Review:	Exempt Determination
Project Title:	Predictors of an Effective Performance Measurement System: Evidence from Municipalities in Turkey
Investigator:	Sedat Eliuz
IRB Number:	SBE-14-10358
Funding Agency:	
Grant Title:	
Research ID:	N/A

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

IRB Coordinator

APPENDIX D: CORRELATION MATRICES

Tablo 32 *Correlation Matrix for Organizational Support*

		OS1	OS2	OS3	OS4	OS5	OS6	OS7
OS1	Correlation Coefficient	1,000						
	Sig. (2-tailed)	.						
	N	428						
OS2	Correlation Coefficient	,768**	1,000					
	Sig. (2-tailed)	,000	.					
	N	428	428					
OS3	Correlation Coefficient	,645**	,763**	1,000				
	Sig. (2-tailed)	,000	,000	.				
	N	428	428	428				
OS4	Correlation Coefficient	,444**	,540**	,650**	1,000			
	Sig. (2-tailed)	,000	,000	,000	.			
	N	428	428	428	428			
OS5	Correlation Coefficient	,494**	,482**	,462**	,460**	1,000		
	Sig. (2-tailed)	,000	,000	,000	,000	.		
	N	428	428	428	428	428		
OS6	Correlation Coefficient	,598**	,516**	,458**	,409**	,698**	1,000	
	Sig. (2-tailed)	,000	,000	,000	,000	,000	.	
	N	428	428	428	428	428	428	
OS7	Correlation Coefficient	,502**	,565**	,511**	,531**	,688**	,738**	1,000
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	.
	N	428	428	428	428	428	428	428

**. Correlation is significant at the 0.01 level (2-tailed).

Tablo 33 *Correlation Matrix for External Support*

		ES1	ES2	ES3	ES4	ES5
ES1	Correlation Coefficient	1,000				
	Sig. (2-tailed)	.				
	N	428				
ES2	Correlation Coefficient	,791**	1,000			
	Sig. (2-tailed)	,000	.			
	N	428	428			
ES3	Correlation Coefficient	,602**	,667**	1,000		
	Sig. (2-tailed)	,000	,000	.		
	N	428	428	428		
ES4	Correlation Coefficient	,510**	,603**	,647**	1,000	
	Sig. (2-tailed)	,000	,000	,000	.	
	N	428	428	428	428	
ES5	Correlation Coefficient	,454**	,540**	,531**	,700**	1,000
	Sig. (2-tailed)	,000	,000	,000	,000	.
	N	428	428	428	428	428

**. Correlation is significant at the 0.01 level (2-tailed).

Tablo 34 *Correlation Matrix for Technical Capacity*

		TC1	TC2	TC3	TC4	TC5	TC6
TC1	Correlation Coefficient	1,000					
	Sig. (2-tailed)	.					
	N	428					
TC2	Correlation Coefficient	,685**	1,000				
	Sig. (2-tailed)	,000	.				
	N	428	428				
TC3	Correlation Coefficient	,817**	,672**	1,000			
	Sig. (2-tailed)	,000	,000	.			
	N	428	428	428			
TC4	Correlation Coefficient	,783**	,650**	,847**	1,000		
	Sig. (2-tailed)	,000	,000	,000	.		
	N	428	428	428	428		
TC5	Correlation Coefficient	,761**	,625**	,841**	,890**	1,000	
	Sig. (2-tailed)	,000	,000	,000	,000	.	
	N	428	428	428	428	428	
TC6	Correlation Coefficient	,667**	,573**	,683**	,713**	,724**	1,000
	Sig. (2-tailed)	,000	,000	,000	,000	,000	.
	N	428	428	428	428	428	428

**. Correlation is significant at the 0.01 level (2-tailed).

Tablo 35 *Correlation Matrix for Validity*

		V1	V2	V3	V4	V5	V6
V1	Correlation Coefficient	1,000					
	Sig. (2-tailed)	.					
	N	428					
V2	Correlation Coefficient	,906**	1,000				
	Sig. (2-tailed)	,000	.				
	N	428	428				
V3	Correlation Coefficient	,784**	,801**	1,000			
	Sig. (2-tailed)	,000	,000	.			
	N	428	428	428			
NL4	Correlation Coefficient	,751**	,761**	,701**	1,000		
	Sig. (2-tailed)	,000	,000	,000	.		
	N	428	428	428	428		
V5	Correlation Coefficient	,781**	,753**	,768**	,816**	1,000	
	Sig. (2-tailed)	,000	,000	,000	,000	.	
	N	428	428	428	428	428	
V6	Correlation Coefficient	,773**	,759**	,720**	,778**	,825**	1,000
	Sig. (2-tailed)	,000	,000	,000	,000	,000	.
	N	428	428	428	428	428	428

**. Correlation is significant at the 0.01 level (2-tailed).

Tablo 36 *Correlation Matrix for Legitimacy*

		L1	L2	L3	L4	L5	L6
L1	Correlation Coefficient	1,000					
	Sig. (2-tailed)	.					
	N	428					
L2	Correlation Coefficient	,595**	1,000				
	Sig. (2-tailed)	,000	.				
	N	428	428				
L3	Correlation Coefficient	,655**	,617**	1,000			
	Sig. (2-tailed)	,000	,000	.			
	N	428	428	428			
L4	Correlation Coefficient	,709**	,495**	,686**	1,000		
	Sig. (2-tailed)	,000	,000	,000	.		
	N	428	428	428	428		
L5	Correlation Coefficient	,681**	,505**	,637**	,802**	1,000	
	Sig. (2-tailed)	,000	,000	,000	,000	.	
	N	428	428	428	428	428	
L6	Correlation Coefficient	,612**	,548**	,594**	,686**	,744**	1,000
	Sig. (2-tailed)	,000	,000	,000	,000	,000	.
	N	428	428	428	428	428	428

**. Correlation is significant at the 0.01 level (2-tailed).

Tablo 37 *Correlation Matrix for Functionality*

		F1	F2	F3	F4	F5
F1	Correlation Coefficient	1,000				
	Sig. (2-tailed)	.				
	N	428				
F2	Correlation Coefficient	,838**	1,000			
	Sig. (2-tailed)	,000	.			
	N	428	428			
F3	Correlation Coefficient	,801**	,823**	1,000		
	Sig. (2-tailed)	,000	,000	.		
	N	428	428	428		
F4	Correlation Coefficient	,799**	,813**	,835**	1,000	
	Sig. (2-tailed)	,000	,000	,000	.	
	N	428	428	428	428	
F5	Correlation Coefficient	,783**	,793**	,810**	,840**	1,000
	Sig. (2-tailed)	,000	,000	,000	,000	.
	N	428	428	428	428	428

**. Correlation is significant at the 0.01 level (2-tailed).

Tablo 38 *Correlation Matrix for Effectiveness of Performance Measurement*

		E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
E1	Correlation C.	1,000									
	Sig. (2-tailed)	.									
	N	428									
E2	Correlation C.	,921**	1,000								
	Sig. (2-tailed)	,000	.								
	N	428	428								
E3	Correlation C.	,831**	,841**	1,000							
	Sig. (2-tailed)	,000	,000	.							
	N	428	428	428							
E4	Correlation C.	,838**	,851**	,790**	1,000						
	Sig. (2-tailed)	,000	,000	,000	.						
	N	428	428	428	428						
E5	Correlation C.	,774**	,781**	,754**	,772**	1,000					
	Sig. (2-tailed)	,000	,000	,000	,000	.					
	N	428	428	428	428	428					
E6	Correlation C.	,723**	,717**	,671**	,702**	,773**	1,000				
	Sig. (2-tailed)	,000	,000	,000	,000	,000	.				
	N	428	428	428	428	428	428				
E7	Correlation C.	,791**	,796**	,763**	,783**	,750**	,741**	1,000			
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	.			
	N	428	428	428	428	428	428	428			
E8	Correlation C.	,764**	,799**	,734**	,775**	,741**	,731**	,885**	1,000		
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	.		
	N	428	428	428	428	428	428	428	428		
E9	Correlation C.	,811**	,829**	,742**	,772**	,738**	,725**	,829**	,848**	1,000	
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	.	
	N	428	428	428	428	428	428	428	428	428	
E10	Correlation C.	,763**	,762**	,748**	,761**	,697**	,723**	,789**	,763**	,807**	1,000
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	.
	N	428	428	428	428	428	428	428	428	428	428

**. Correlation is significant at the 0.01 level (2-tailed).

APPENDIX E: PRELIMINARY CONFIRMATORY FACTOR ANALYSIS RESULTS

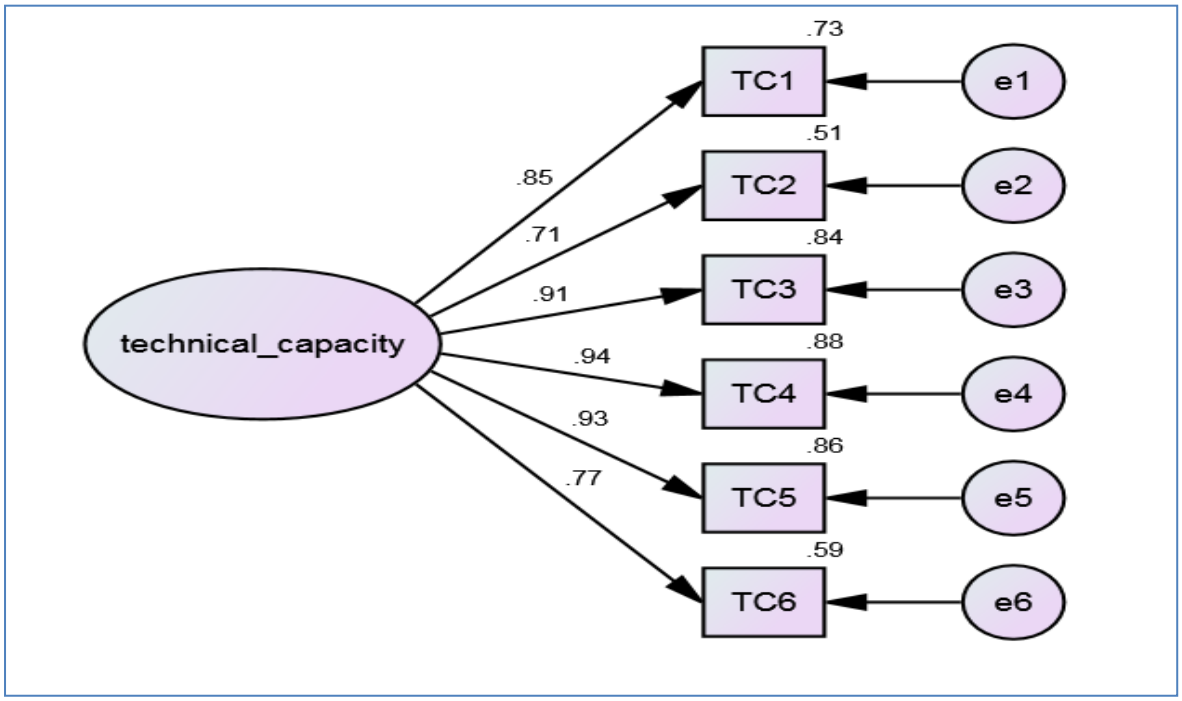

Figure 30. Preliminary Confirmatory Factor Analysis Results for Technical Capacity

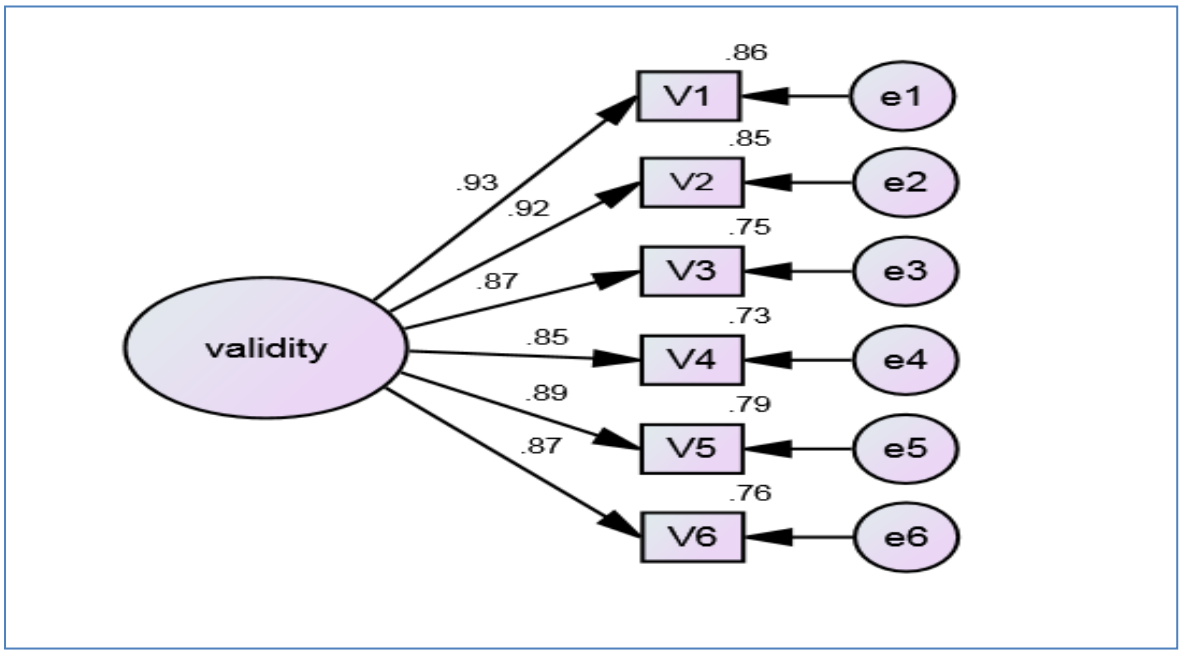

Figure 31. Preliminary Confirmatory Factor Analysis Results for Validity

REFERENCES

Abramson, M., Bruel, J., & Kamensky, J. (2006). *Six trends transforming government*. Washington, DC: IBM Center for the Business of Government.

Agiro, A. T. (2011). *Determinants of productivity in hospital-based rural health clinics: a growth curve modeling approach* (Unpublished doctoral dissertation). University of Central Florida, Orlando, Fl.

Ammons, D.N. (2002). Performance measurement and managerial thinking. *Public Performance & Management Review,* 25(4), 344–347.

Ammons, D.N. (2013). Signs of performance measurement progress among prominent city governments. *Public Performance & Management Review*, 36(4), 507-528.

Ammons, D. N., & Rivenbark, W. C. (2008). Factors influencing the use of performance data to improve municipal services: Evidence from the North Carolina Benchmarking Project. *Public Administration Review*, 68(2), 304-318.

Ates, H. & Cetin, D. (2004). Kamuda performans yönetimi ve performansa dayalı bütçe [Performance management and performance-based budgeting in public administration]. In M. Acar & H. Ozgur (Eds.), *Çağdaş Kamu Yönetimi II* [Contemporary Public Administration II] (pp. 255-284), Ankara, Turkey: Nobel.

Babbie, E. R. (2013). *The Practice of social research* (13th ed.). Belmont, CA : Wadsworth Cengage Learning.

Bayraktar, S.U. (2007). Turkish municipalities: Reconsidering local democracy beyond administrative autonomy, *European Journal of Turkish Studies [online]*, Retrieved from http://ejts.revues.org/1103

Behn, R. D. (1995). The big questions of public management. *Public Administration Review*, 55(4), 313-324.

Behn, R.D. (2001). *Rethinking democratic accountability.* Washington, DC: Brookings Institution Press.

Behn, R.D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review*, 63(5), 586-606.

Bentler, P. M. & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods and Research*, 16(1), 78-117.

Berkman, A. U. & Heper, M. (2002). Political Dynamics and Administrative Reform in Turkey. In A. Farazmand (Ed.) *Administrative reform in developing nations* (pp. 151-162). Westport, CT: Praeger Pub Text.

Berman, E. & Wang, X. (2000). Performance measurement in U.S. counties: Capacity for reform. *Public Administration Review*, 60(5), 409-420.

Bititci, U. S., Turner, T., & Begemann, C. (2000). Dynamics of performance measurement systems. *International Journal of Operations & Production Management*, 20(6), 692-704.

Boomsma, A., & Hoogland, J.J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. du Toit & D. Sörbom (Eds.), *Structural equation models: Present and future. A Festschrift in honor of Karl Jöreskog* (pp. 139–168). Lincolnwood, IL: Scientific Software International.

Bouckaert, G. (1993). Measurement and meaningful management. *Public Productivity and Management Review*,17(1), 31-44.

Boyer, S., & Martin, L.L. (2012). The Florida Benchmarking Consortium: A local government performance measurement & benchmarking network. *Public Performance & Management Review,* 36(1), 124-137.

Broad, M. L. (2006). Improving performance in complex organizations. *Industrial & Commercial Training*, 38(6/7), 322-329. doi:10.1108/00197850610685833.

Burke, W.W., & Litwin, G. H. (1992). A causal model of organizational performance and change. *Journal of Management*, 18(3), 523-545.

Byrne, B.M. (2010). *Structural equation modeling with AMOS.* New York, NY: LEA Publisher.

Celenk, A. (2009). Europeanization and administrative reform: The case of Turkey. *Mediterranean Politics*, 14(1), 41-60. doi:10.1080/13629390902747392

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.

Creswell, J.W. (2009). *Research design: Qualitative, quantitative, and mixed methods and approaches* (3rd ed.). Los Angeles, CA: SAGE Publications.

da Cruz, N., & Marques, R. (2014). Revisiting the determinants of local government performance. *Omega*, 4491-103. doi:10.1016/j.omega.2013.09.002

Dalehite, E. G. (2008). Determinants of performance measurement: an investigation into the decision to conduct citizen surveys. *Public Administration Review*, (5), 891-907.

de Lancer Julnes, P. L. & Holzer, M. (2001) Promoting the utilization of performance measures in public organizations: An empirical study of factors affecting adopting and implementation. *Public Administration Review*. 61(6), 693-708.

de Lancer Julnes, P., & Holzer, M. (2008). *Performance measurement: Building theory, improving practice*. New York, NY: M.E. Sharpe.

Denhardt, R.B., & Denhardt, J.V. (1999). *Leadership for change: Case studies in American local government*. Washington, DC: IBM Center for the Business of Government.

Dillman, D.A., Smyth, J.D., & Christian, L.M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). Haboken, NJ: Wiley.

Dixon, J., Nanni, A. J., & Vollmann, T. E. (1990). *The new performance challenge: Measuring operations for world-class companies.* Homewood, Ill.: Dow Jones-Irwin.

Donabedian, A. (1988). The quality of care: How can it be assessed? *Journal of the American Medical Association*, 260(12), 1743–1748.

Donaldson, L. (2001). *The contingency theory of organizations*. Thousand Oaks, CA: Sage Publications.

Drazin, R., & Van De Ven, A. H. (1985). Alternative forms of fit in contingency theory. *Administrative Science Quarterly*, 30(4), 514–539.

Edwards, D., & Thomas, J. (2005). Developing a municipal performance-measurement system: Reflections on the Atlanta Dashboard. *Public Administration Review*, 65(3), 369-376.

Fernandez, S., & Rainey, H. G. (2006). Managing successful organizational change in the public sector. *Public Administration Review*, *66*(2), 168-176. doi:10.1111/j.1540-6210.2006.00570.x

Flood, A. B., Zinn, J. S., & Scott, W. R. (2006). Organizational performance: Managing for efficiency and effectiveness. In S.M. Shortell & A.D. Kaluzny (Eds.), *Health care*

management: Organization design and behavior (5th ed., pp. 415–457). Clifton Park, NY: Thomson Delmar Learning.

Folz, D. H., Abdelrazek, R., & Chung, Y. (2009). The adoption, use, and impacts of performance measures in medium-size cities: Progress toward performance management. *Public Performance & Management Review,* 33(1), 63-87.

Fryer, K., Antony, J., & Ogden, S. (2009). Performance management in the public sector. *International Journal of Public Sector Management*, 22(6), 478-498. doi:10.1108/09513550910982850

Garson, G. D. (2012). Structural equation modeling. Retrieved from http://www.statisticalassociates.com/sem.pdf

Garver, M. S., and Mentzer, J. T. (1999). Logistics research methods: Employing structural equation modeling to test for construct validity. *Journal of Business Logistics*, 20, 33-58.

George, D., & Mallery, P. (2007). *SPSS for windows step by step: a simple guide and reference, 14.0 update*. Boston, MA: Pearson/A and B.

Goltz, J. (2006). *Police organizational performance in the state of Florida: Confirmatory analysis of the relationship of the environment and design structure to performance* (Unpublished doctoral dissertation). University of Central Florida, Orlando, Fl.

Greiling, D. (2006). Performance measurement: A remedy for increasing the efficiency of public services?. *International Journal of Productivity and Performance Management*, 55(6), 448-465. doi:10.1108/17410400610682488

Guney, A., & Celenk, A. (2010). Europeanization and the dilemma of decentralization: Centre-local relations in Turkey. *Journal of Balkan & Near Eastern Studies*, 12(3), 241-257. doi:10.1080/19448953.2010.506810

Guven, A. (2012) *Belediyelerde performans denetimi ve algılanması üzerine bir araştırma* [A resarch on performance audit and perception in municipalities]. (Unpublished doctoral dissertation). University of Istanbul, Istanbul, Turkey.

Häggroth, S. (2013). Leading change by including stakeholders. *Public Administration Review*, 73(4), 550. doi:10.1111/puar.12095

Hatry, H.P. (2006). *Performance measurement: Getting results* (2nd edition). Washington, DC: Urban Institute Press.

Hill, C.J., & Lynn, L.E. (2009). *Public management: A three dimensional approach*. Washington DC: Congressional Quarterly Press.

Ho, A.T. (2002). Reinventing local governments and the e-government initiative. *Public Administration Review*. 62(4), 434-444.

Ho, A.T. (2006). Accounting for the value of performance measurement from the perspective of Midwestern mayors. *Journal of Public Administration Research and Theory*. 16(2), 217-238.

Hoe, S. L. (2008). Issues and procedures in adopting structural equation modeling technique. *Journal of Applied Quantitative Methods*, 3(1), 76-83.

Holzer, M. & Yang, K. (2004). Performance measurement and improvement: An assessment of the state of the art. *International Review of Administrative Sciences*. 70(1), 15-31.

Hoontis, P., & Kim, T. (2012). Antecedents to municipal performance measurement implementation: a case study of two New Jersey municipalities. *Public Performance and Management Review*, 36(1), 158-173. doi:10.2753/PMR1530-9576360108

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.

Hubbard, D.W. (2010). *How to measure anything: finding the value of "intangibles" in business* (2nd ed.). Hoboken, NJ: Wiley.

Hughes, O.E. (2012). *Public management and administration: An introduction* (4th ed.). London, UK: Palgrave.

Ingraham, P.W., Donahue, A., & Joyce, P.G. (2003). *Government performance: Why management matters*. Baltimore, MD: Johns Hopkins University Press.

Johnsen, Å. (2005). What does 25 years of experience tell us about the state of performance measurement in public policy and management?. *Public Money & Management*, 25(1), 9-17. doi:10.1111/j.1467-9302.2005.00445.x

Kaplan, D. (1995). Statistical power in structural equation modeling. In R.H. Hoyle, (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 100-117). Thousand Oaks, CA: Sage Publications, Inc.

Kaplan, R.S., & Norton, D.P. (1992). The Balanced Scorecard--Measures that drive performance. *Harvard Business Review*, *70*(1), 71-79.

Kapucu, N. (2010). *Governance reforms: Comparative perspectives*. Ankara, Turkey: International Strategic Research Organization (ISRO) Publications.

Kapucu, N. & Palabiyik, H. (2008). *Turkish public administration: From tradition to modern age.* Ankara, Turkey: International Strategic Research Organization (ISRO) Publications.

Kapucu, N., Volkov, I., & Wang, X. (2011). Best practices of best performers: Evidence from the Florida Sterling Council award winners. *Public Performance and Management Review*, 34(3), 397-419.

Karasu, M., & Demir, M. (2012). Stratejik planlama, performans esaslı bütçeleme ve büyükşehir belediyelerindeki uygulamalar [Strategic planning, performance budgeting, and the practices in metropolitan municipalities]. *Suleyman Demirel University Journal of Faculty of Economics & Administrative Sciences*, 17(3), 339-349.

Kettl, D. (1994). *Reinventing government? Appraising the National Performance Review*. Washington, DC: The Brookings Institution.

Kline, R. B. (2011). Principles and practice of structural equation modeling (3rd ed.). New York, NY: Guilford Press.

Klovienė, L., & Gimžauskienė, E. (2008). The role of institutional factors on changes of performance measurement system. *Economics & Management*, 22-29.

Kosecik, M., Kapucu, N., & Sezer, Y. (2003). Becoming an effective public manager in the global world: District governors in Turkey. *D. E. University Economics and Administrative Sciences Journal,* 18(1), 105-129.

Koseoglu, O. (2005). Belediyelerde performans yönetimi [Performance management in municipalities]. *Turk Idare Dergisi* [Journal of Turkish Administration], 447, 211-234.

Lei, P., & Wu, Q. (2007). Introduction to structural equation modeling: Issues and practical considerations. *Educational Measurement: Issues and Practice*, 3, 33. doi:10.1111/j.1745-3992.2007.00099.x

Lin, B., & Wan, T. (1999). Analysis of integrated healthcare networks' performance: a contingency-strategic management perspective. *Journal of Medical Systems*, 23(6), 467-485.

Maguire, E. R. (2003). *Organizational structure in American police agencies: Context, complexity, and control*. Albany, NY: State University of New York Press.

Marathe, S. (2006). Benchmarking community health centers' efficiency: Multivariate analysis (Unpublished doctoral dissertation). University of Central Florida, Orlando, Fl.

Mark, B., Salyer, J., & Wan, T. (2003). Professional nursing practice: Impact on organizational and patient outcomes. *The Journal of Nursing Administration*, 33(4), 224-234.

Martin, L.L. (2002). *Making performance-based contracting reform: What the federal government can learn from state and local governments*. New Ways to Manage Series. Washington, DC: IBM Endowment for the Business of Government.

Martin, L. L., & Frahm, K. K. (2010). The changing nature of accountability in administrative practice. *Journal of Sociology and Social Welfare*, 37(1), 137-148.

Martin, L. L., & Kettner, P. M. (2010). *Measuring the performance of human service programs*. Los Angeles, CA: Sage.

Mausolff, C., & Spence, J. (2008). Performance measurement and program effectiveness: A structural equation modeling approach. *International Journal of Public Administration*, 31(6), 595-615. doi:10.1080/01900690701640929

McDonald, R. P., & Ho, M.H. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7(1), 64-82.

Melkers, J., & Willoughby, K. (2005). Models of performance-measurement use in local governments: Understanding budgeting, communication, and lasting effects. *Public* Administration Review, 65(2), 180-190. doi:10.1111/j.1540-6210.2005.00443.x

Meyers, L.S., Gamst, G., & Guarino, A.J. (2013). *Applied multivariate research: Design and interpretation* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Modell, S. (2001). Performance measurement and institutional processes: A study of managerial responses to public sector reform. *Management Accounting Research*, 12(4), 437-464.

Morgan, G., Leech, N., Gloekner, G., & Barrett, K.C. (2005). *SPSS for introductory statistics: Use and interpretation*. Mahwah, NJ: Erlbaum.

Moynihan, D.P. (2005). Why and how do state governments adopt and implement "Managing for Results" reforms? *Journal of Public Administration Research and Theory*. 15(2), 219-243.

Moynihan, D. P. (2006). Managing for results in state government: Evaluating a decade of reform. *Public Administration Review*. 66(1), 77-89.

Moynihan, D. P. (2008). *The dynamics of performance management: Constructing information and reform*. Washington, DC: Georgetown University Press.

Nalbandian, J., O'Neill, R., Wilkes, J. M., & Kaufman, A. (2013). Contemporary challenges in local government: Evolving roles and responsibilities, structures, and processes. *Public Administration Review*, 73(4), 567-574. doi:10.1111/puar.12059

OECD. (1998). *Public management reform and economic and social development.* Paris, France: OECD.

Oliver, C. (1991). Strategic responses to institutional processes. *Academy of Management Review*, 16(1), 145-179. doi:10.5465/AMR.1991.4279002.

Ozcan, G. B, & Turunc, H. (2008). The politics of administrative decentralization in Turkey since 1980. In J. Killian, & N. Eklund (Eds.), *Handbook of administrative reform: An international perspective* (pp. 177-196). New York, NY: Taylor & Francis Group.

Perrin, B. (1998). Effective use and misuse of performance measurement. *American Journal of Evaluation*, 19(3), 367-379.

Poister, T. H., & Streib, G. (1999). Performance measurement in municipal government: Assessing the state of the practice. *Public Administration Review*, 59(4), 325-335.

Pollanen, R. M. (2005). Performance measurement in municipalities: Empirical evidence in Canadian context. *International Journal of Public Sector Management*, 18(1), 4-24.

Radin, B. (2006). *Challenging the performance movement: Accountability, complexity, and democratic values*. Washington, DC: Georgetown University Press.

Rejc, A. (2004). Toward contingency theory of performance measurement. *Journal for East European Management Studies*, 9(3), 243-264.

Rivenbark, W. C., & Kelly, J. M. (2003). Management innovation in smaller municipal government. *State and Local Government Review*, 35(3), 196–205.

Rosenberg, R. (1998). *Breaking out the change trap: A practical guide for organizational change.* Raleigh, NC: Banbury Press.

Salazar, J. I., & Martínez, C. L. (2013). Adoption of performance measurement: Evidence of Mexican municipalities. *International Journal of Public Administration*, 36(11), 754-767. doi:10.1080/01900692.2013.794833

Sanger, M. B (2008). From measurement to management: Breaking through the barriers to state and local performance. *Public Administration Review,* 68, S70-S85.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23-74.

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). New York, NY: Routledge/Taylor & Francis Group.

Scott, W. (2003). *Organizations: Rational, natural, and open systems*. Upper Saddle River, NJ: Prentice Hall.

Scott, R. & Davis, G. (2007). *Organizations and organizing: Rational, natural and open systems perspectives*. Upper Saddle River, NJ: Prentice Hall.

Sezen, S. (2011). International versus domestic explanations of administrative reforms: The case of Turkey. *International Review of Administrative Sciences*, 77(2), 322-346.

Sivo, S., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for "Optimal" cutoff properties: Fit index criteria in structural equation modeling. *Journal of Experimental Education,* 74(3), 267-288.

Sozen, S., & Shaw, I. (2002). The international applicability of 'new' public management: Lessons from Turkey. *International Journal of Public Sector Management*, 15(6), 475-486.

Streib, G., & Poister, T. (1999). Assessing the validity, legitimacy and functionality of performance measurement systems in municipal governments. *American Review of Public Administration*, 29(2), 107-123.

Taylor, J. (2006). Performance measurement in Australian and Hong Kong government departments. *Public Performance & Management Review*, 29(3), 334-357.

Taylor, J. (2011). Factors influencing the use of performance information for decision making in Australian state agencies. *Public Administration*, 89(4), 1316-1334. doi:10.1111/j.1467-9299.2011.02008.x

The U.S. Government Accountability Office. (2011, May*). Performance Measurement and Evaluation: Definitions and Relations,* (Publication No. GAO-11-646SP). Retrieved from http://www.gao.gov

Thompson, B. (2000). Ten Commandments of structural equation modeling. In L. G. Grimm, P. R. Yarnold (Eds.), *Reading and understanding MORE multivariate statistics* (pp. 261-283). Washington, DC: American Psychological Association.

Turkish Ministry of Interior. (2014a). Field Administrations. Retrieved from: https://www.e-icisleri.gov.tr/Anasayfa/MulkiIdariBolumleri.aspx

Turkish Ministry of Interior. (2014b). Belediyeler [Municipalities]. Retrieved from: http://www.migm.gov.tr/IstatistikiBilgiler.aspx?DetayId=1

Ullman, J. B. (2006). Structural equation modeling: Reviewing the basics and moving forward. *Journal of Personality Assessment*, 87(1), 35-50. doi:10.1207/s15327752jpa8701_03

Ullman, J. B. (2007). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (5th ed., pp. 676–780). New York, NY: HarperCollins.

United Nations Public Administration Network (n.d.) *Assessment of Public Administration in Turkey*. Retrieved from http://unpan1.un.org/intradoc/groups/public/documents/un/unpan031588.pdf

Van de Welle, S. (2008). Comparing the performance of national public sectors: Conceptual problems. *International Journal of Productivity & Performance Management*, 57(4), 329-338.

Van Thiel, S., & Leeuw, F. L. (2002). The performance paradox in the public sector. *Public Performance & Management Review*, 25(3), 267-281.

Wan, T. (1995). *Analysis and evaluation of health care systems: An integrated approach to managerial decision making*. Baltimore, MD: Health Professions Press.

Wan, T. T. H. (2002). *Evidence-Based health care management*. Boston, MA: Kluwer Academic Publishers.

Wang, X. (2002). Performance measurement impact: A study of U.S. local governments. *Public Performance and Management Review*. 26(1), 26-43.

Wang, X. (2010). *Performance analysis for public and nonprofit organizations.* Sudbury, MA: Jones and Bartlett Publishers.

Wang, X., & Berman, E. (2001). Hypotheses about performance measurement in counties: Findings from a survey. *Journal of Public Administration Research & Theory*, 11(3), 403-428.

Wang, X., Hawkins, C. V., Lebredo, N., & Berman, E. M. (2012). Capacity to sustain sustainability: A study of U.S. cities. *Public Administration Review*, 72(6), 841-853. doi:10.111/j.1540-6210.2012.02566.x.

Weill, P., & Olson, M. H. (1989). An assessment of the contingency theory of management information systems. *Journal of Management Information Systems*, 6(1), 59-85.

Williams, D. W. (2003). Measuring government in the early twentieth century. *Public Administration Review*, *63*(6), 643-659. doi:10.1111/1540-6210.00329.

Yang, B. (2002). Meta-Analysis research and theory building. *Advances in developing human resources,* 4(3), 296-316. doi: 10.1177/1523422302043005.

Yang, K. & Holzer, M. (2006). The performance-trust link: Implications for performance measurement. *Public Administration Review*. 66(1), 114-126.

Yang, K., & Hsieh, J. (2007). Managerial effectiveness of government performance measurement: Testing a middle-range model. *Public Administration Review*, 67(5), 861-879.

Yang, Y., & Wu, J, (2013). Are the "Bigger Fish" caught? China's experience of engaging citizens in performance measurement system. *Public Administration Quarterly*, 37(2), 144-181.

Zamesnik, K. A. (2012). Local government performance measurement: A commentary. Retrieved from http://repositories.lib.utexas.edu/bitstream/handle/2152/ETD-UT-2012-08-6342/ZAMESNIK-MASTERS-REPORT.pdf?sequence=1

Zinn, J. S., & Mor, V. (1998). Organizational structure and the delivery of primary care to older Americans. *Health Services Research*, 33(2), 354-380.