Between Abandonment and Adoption:

The Value of E-valu-ation for Turkey's Educational Decision Making

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY
OF MINNESOTA
BY

Hanife Cakici

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jean A. King, Adviser
David Chapman, Adviser

June, 2014

Acknowledgments

This endeavor is solely dedicated to my mother, Bediya Çakıcı, whose life story is worth a thousand dissertations. This is a tribute to her dreams that I have embodied as mine. Her unending encouragement to dare dream beyond the border of a small, Anatolian village to the skyscrapers of the American Dream and to venture into three different continents deserves the deepest appreciation. I am truly and deeply indebted to her mere existence in this world. *Annem; nurlu ellerinle ve dualarınla yön verdiğin hayatım, dünyanın başka yerlerinde başka kız çocuklarının hayatlarına dokunmak için çoktan yola koyuldu.*

And I am truly grateful to my father, Mükerrem Çakıcı, and three siblings, Mustafa, Hilal and Zeliha Çakıcı, whose life trajectories crystalized Turkey's long-term image of a *bridge*, standing right at the cross section of tradition and modernity, religion and science, East and West, South and North, summer and winter. My heartfelt thanks to my nuclear family – the wonder of the wonders – for challenging me in unprecedented ways and keeping their horizons larger than any other family that I know.

A special thanks goes to my advisor, Dr. Jean King, the Queen of Evaluation! My life is forever on a different path ever since I attended Dr. King's class of *Principles and Methods of Program Evaluation* in 2010. Her contagious passion for evaluation, unwavering support and mentorship, and much needed sense of humor in academia is duly appreciated. Every glimmer of hope Dr. King has injected into my academic life over the years in the face of uncertainty spilled over to my entire career.

ending speeches full with evaluation jargon, namely Jeffrey Paul Engell – my sole

inspiration – and my beloved Colorado family, Nora LeRoux Radtke, David Alexander

Greenwood Sanchez, Candi Walz, and Erin Randall, and finally Anna Swanson and my

beloved Minnesota family – this is my humble way of "thanksgiving."

I would also like to acknowledge the Bohemian coffee shops of the Twin Cities –

specifically the Purple Onion Café, Bordertown Coffee, Espresso Royale Caffe,

Starbucks, Mapps Coffee and Tea, and Overflow Espresso Café – for keeping me

caffeinated and entertained as I endured the experience of graduate school.

Namaste, CorePower Yoga for maintaining my sanity and reminding me of

setting dedications while making meaning of this study.

Last but never the least, a shout out to bittersweet Minnesota winters for leaving

me with nothing but an unsolicited desire to sit down and just write my dissertation.

Thank you, Minnesota, for being *nice* to me!

And my beloved land of the Turks, I will never have the right words to describe

my unstoppable, untainted, passionate love for you. *Ne Mutlu Türk'üm Diyene!*

Abstract

This dissertation examines the value of evaluation for Turkey's educational decision making. Relying on individual interviews and document review, the study analyzes how key stakeholders – government officials, academics, and civil society representatives – and governmental acts envision and portray the role and utility of evaluation for public decision making specifically in the education sector. Drawing on social science literature from the fields of evaluation, comparative education, public policy, and international development, this study addresses the need to decolonize the concept and practice of evaluation, as this trans-disciplinary field is rapidly cutting across geographic, historic, social, and cultural borders. This study revisits the origins of evaluation practice in the global Northern context, traces its expansion into the global South across a number of sites, and argues that context matters in transferring, borrowing, negotiating, establishing, practicing, and using the concept and practice of evaluation.

Evidence for this study's conclusions comes from Turkey's relatively immature history with evaluation in the education decision domain. Motivated by the desire to become one of the top ten largest economies in the world by 2023, Turkey's rapid development underlined educational achievement and growth as the roadmap. This quest necessitated a specific form of educational governance and decision making driven by the principles of effectiveness and efficiency. At the heart of these principles, monitoring and evaluation (M&E) has long lain as a tool of accountability, learning, and improvement, in which Turkey's entire public administrative culture has historically lagged behind. In response to this immaturity, supranational authorities and international donors have

provided financial and technical impetus for locating M&E systems, practice, and information in the Turkish education decision domain. Coupled with the country's official drive for modernization, international actors, to a great extent, paved the way to legal arrangements for streamlining evaluation. Specifically, the Green Paper published after the European Union's "*Strengthening the Capacity of the Ministry of National Education Project*" later became the conceptual foundation for Decree No. 652 that helped establish M&E units at the Ministry for the first time in Turkey's educational history.

Despite all these efforts, the study reveals that evaluation remains as a new concept that is closely associated with quantification, performance-based budgeting, and compliance. Evaluation's value mostly resides in its symbolic representation of modern norms of governance to which Turkey eagerly wants to commit. Yet, reported confusion about what evaluation really entails, as a concept and as a field of practice, is paired with highly centralized and politically polarized educational governance, all together situating evaluation in foster care in Turkey: it is neither fully adopted, nor is it completely abandoned.

Adopting a constructivist-critical outlook on the role of evaluation in the global South, this study endeavors to locate this field of practice in the broader context of international development with its negotiated margins, borders, and struggles. By suggesting that evaluation is a marker of a country's quest to modernize and Westernize, this study sheds light on the direction of cross-cultural expansion of the field of evaluation.

Table of Contents

List of Tables

List of Figures

Chapter I: Introduction

Problem Statement

In 2012, The Republic of Turkey's Ministry of National Education (MoNE) launched the "School Milk Project" in cooperation with the Ministry of Health and the Ministry of Food, Agriculture, and Livestock. The aim of this project was to support students' engagement and learning in school by improving their nutritional habits. A few months after the project started, hundreds of students had been admitted into hospitals because of spoiled milk and food poisoning. Major newspapers and network news in Turkey covered these cases, and the incidents were considered a public disgrace, all the more humbling because Turkey was considered to be on the road to European Union (EU) membership. This incident revived the longstanding debate about governmental decision making and implementation of public programs and policies in Turkey. Lurking behind this debate were three common questions: Did educational authorities do the right thing in designing this project? Did they implement it correctly? What could they have done better?

These questions are linked to a broader debate about the role and utility of evaluation in decision making by Turkish educational officials. Scholars suggest that evaluations play a strategic role in decision-making processes by generating a continuous flow of solid information about the merit, shortcomings, and outcomes of public programs and policies, contributing to their effectiveness and betterment (Weiss, 1977; Patton, 2012; Mark & Henry, 2004; Fitzpatrick, Sanders, & Worthen, 2004). While the notion that evaluations are essential in any society is nowadays commonplace in the mainstream evaluation literature, available evidence suggests that formal program and

policy evaluations are rare in Turkey (Gür & Çelik, 2009; ERI, 2010a, 2010b; Russon & Russon, 2000). Aydagül (2013) once observed:

> An important conclusion of recent analyses is the lack of assessments on the impact and cost-effectiveness of public policies, programmes, and projects implemented to achieve equity and quality in education. The policy life cycle often falls short of an empirical, objective, and rigorous evaluation phase, which impedes further policymaking cycles. Lessons learned from previous experiences should constitute a rich learning course for the MoNE, but often do not. (pp. 405-406)

As a result, it is speculated that, as a country that aims to cope with competitive pressure within the EU and global knowledge economy, Turkey may frequently lose the opportunity to improve educational policies and programs in compliance with the international standards of educational quality (ERI, 2013a). To this end, Turkish scholars and practitioners have increased their calls for home evaluations of national programs and policies grounded in Turkey's social, economic, and cultural context (TEPAV, 2013; Aydagül, 2008; Şişman, 2011).

The historical development of evaluation culture in Korea and Brazil – with comparable levels of development to Turkey – may suggest the potential for utilizing evaluation as a decision-making tool in Turkey to provide useful information about policies' effectiveness in improving education quality and equity. With the launch of "Government for People" in 1998 as a response to severe economic crisis in Asia, Korea has developed government-wide evaluation systems to create and implement national reform packages based on national needs and priorities (Lee, 2002). Despite the insufficient number of evaluators, Korea is praised for conducting several major evaluations, ranging from evaluations of ministries' major programs and policies to meta-evaluations of each institution's policy making and evaluation capacity (Furubo, Rist, &

Sandahl, 2002). Similarly, the evaluation field in Brazil has grown dramatically in recent years with 453 post-graduate evaluation courses and a 90% increase in publicity on government evaluations, taken as indicators of government's desire to foster better programming and budgeting (UNDP, 2011). Turkey's Ninth National Development Plan (2007 – 2013) also praised social and economic developments taking place in Korea and Brazil thanks to their governance reform agenda, and argued that their influence in international decisions will considerably increase in the coming decades (Ministry of Development, 2006). Although the association between increased evaluation activities and better development outcomes is not yet empirically documented, anecdotal evidence suggests a positive relationship (UNDG, 2011; Segone, 2008, 2009).

In contrast to Korea and Brazil, little is known about the implementation and impact of many educational policies, programs, and projects in Turkey, although the continuous improvement of educational services is believed to be of utmost importance to the country's long-term aspirations (ERI, 2010b; Ergüder, 2013). Vision 2023, an official statement of the national goals for the centennial year of the Republic, envisions that Turkey becomes one of the top ten economies in the world with $2 trillion GDP, and $25,000 GDP per capita (Vision 2023, 2013). The 10th Development Plan (2014-2019) reinstates and strengthens Turkey's commitment to "development for and with people," underlining quality education for all as a prerequisite for a strong society (Ministry of Development, 2013). Although the most recent development plan – compared to previous ones – puts a recognizable and intentional emphasis on monitoring and evaluation (M&E) as important elements of good governance, the plan once falls short of explicating how

evaluations will be conducted and utilized, as well as why evaluation could be meaningful for the policy life cycle. The Economic Policy Research Foundation of Turkey (TEPAV) (2013) made a similar observation by arguing that M&E activities indeed foreground the 10th plan's determination for success:

> The benefits of having a state department responsible for monitoring and evaluating the impacts of the policies in the plan are evident. It would be wise to define the function of impact assessment, which is a weak aspect of Turkey's public administration culture, uniquely for this plan and design this together with the monitoring and coordination activities. (p. 27)

In addition to a lack of evaluation of national plans, "The periodical assessment surveys undertaken by the Ministry [of National Education] are underutilized and fail to attract any public attention, thereby detached from the policy-making process," remarked Aydagül (2013, p. 223). In the absence of systematic, formal policy or program evaluations, basic research or/and international studies such as PISA (Program for International Student Assessment), TIMMS (Trends in Mathematics and Science Study), TALIS (Teaching and Learning International Survey) and PIRL (Progress in International Reading Literacy Study) assessments are utilized to remedy the evidence gap in the education policy cycle (see ERI, 2010b; EARGED, 2009, 2010), further contributing to the mismatch between national educational policy design, implementation, and improvement (Gür, 2011).

There is a recent, emergent interest within the civil society and MoNE in strengthening the connection between evaluation and decision making although little is known about the ascribed value of evaluation. First and foremost, in the absence of purposeful and noteworthy governmental effort to conduct and utilize evaluations,

nongovernmental actors were called on to assume responsibility. A lack of knowledge about the impact and shortcomings of educational initiatives motivated the Istanbul Policy Center at Sabancı University, one of the leading research universities in Turkey, to launch the *Education Reform Initiative* (hereafter ERI) in 2003. ERI aims to improve educational decision making and cultivate a new policy-making culture in the country through research, advocacy, and training. This initiative is based on the premise that "decisions are based on data and evaluation and on a transparent and participatory interaction among the state, civil society organizations and citizens" (ERI, 2010b, p. 6). To this end, this initiative facilitates a participatory, democratic public dialogue about educational policies and programs by bringing together representatives from government, civil society, academia, schools, and public and private organizations. The reform emphasizes the importance of informed, evidence-based decision making, best practices, and transparent solutions for alleviating pressing educational problems through its flagship *Education Monitoring Reports* annually published since 2008. In doing so, the Reform hopes to inform decision makers' priorities and practices to help Turkey achieve its long-term aspirations. Although ERI promotes and maintains a discourse for a paradigm shift in educational decision making that is based on and guided by data, the initiative pays insufficient attention to what evaluation can offer to Turkey's education decision domain given the country's lack of history with this concept and practice.

Parallel to ERI, a policy window of opportunity has opened within the Turkish government that positioned evaluation as a decision-making tool in educational programming and planning. In response to increasing demand for effective public

administration specifically by the EU Acquis – the collection of shared rights and

obligations binding all EU member states – as well as the International Monetary Fund

(IMF) stand-by agreements, the Turkish National Grand Assembly enacted Public

Financial Management and Control Law (PFMC) Act No. 5018 in 2003 to require every

public institution to prepare and implement a strategic plan (Ministry of Finance, 2003).

After taking the office on November 2002, the current single-party government declared

its desire to "unleash Turkey's development potential" in the Letter of Intent to IMF by

introducing and implementing a series of stabilization and reform efforts. They promised:

"Our government will particularly focus on a renewed privatization effort, measures to

attract foreign direct investment, fighting corruption, and improving corporate

governance and transparency in the public sector" (Babacan & Serdengeçti, 2003, para.

3), to which Act No. 5018 was proposed as a solution. Although the law strictly

associates evaluation with internal control, audit, and performance-based budgeting

(Yenice, 2006), it has set a precedent in Turkish governmental life by envisioning the

need to install evaluation systems in governmental departments (see European

Commission, 2010). The Article 9 (Ministry of Finance, 2003) explicates the role of

evaluation in policy cycle for the first time in an official policy document:

> Public administrations shall prepare strategic plans in a cooperative manner in
> order to form missions and visions for future within the framework of
> development plans, programs, relevant legislation and basic principles adopted; to
> determine strategic goals and measurable objectives; to measure their
> performances according to predetermined indicators and to monitor and evaluate
> this overall process. (p. 6)

Keeping in line with this directive, MoNE created its first strategic plan in history

as a tool to design, implement, and improve its institutional goals, principles, policies,

and programs (Türk, Yalçın, & Ünsal, 2006). The preface to this first plan authored by the then Minister, Nimet Baş, repeated Turkey's yearning for equal participation in the modern world, hence the inevitability of her change in accordance with global trends in public administration and education systems (Ministry of National Education, 2009). As such, the need for a strategic plan was situated within the context of globalization and modernization that inscribes a standard form of governance highlighting the principles of performance, efficiency, effectiveness, accountability, and transparency promoted and realized by global education actors such as the World Bank and OECD. The following strategic goals clearly indicate that the Ministry set out to build institution-wide M&E systems to improve its decision making (Ministry of National Education, 2009):

> Strategic goal 15.3: Total quality management system will be established, and globally validated evaluation models will be used to monitor educational institutions' performance including schools.
>
> Strategic goal 16.1: Curriculums will be designed and continuously improved based on systematic monitoring and evaluation activities.
>
> Strategic goal 17.4: Monitoring and evaluation units will be established to assess the level and extent to which strategic goals are being accomplished so that corrective measures can be taken.

Nevertheless, empirical evidence is still missing to this date with regards to how M&E is conducted and utilized to inform educational planning and development of subsequent educational strategies (ERI, 2013a). The 209-page long strategic plan dedicated only one page to M&E that fell significantly short of foregrounding evaluation's value for educational planning and programming.

Similarly, the EU's *Capacity Building Support for the Ministry of National Education* (MEBGEP) (2008-2010) created another opportunity for the Ministry to

design and implement effective policies and programs based on evaluation information. The aim of this pre-accession assistance, totaling $4.9 million, was to improve the efficiency and effectiveness of the Turkish education system by developing MoNE's planning, implementation, and monitoring capacity so that educational policies and programs would be harmonized with the EU priorities. One of the central objectives of the grant was to strengthen the capacity of human resources deployed in the education system. This was done by a series of training courses and workshops on topics including data collection, analysis and protection, problem solving and decision making, performance management, monitoring and evaluation, and the use and interpretation of statistics in education (European Commission, 2006a).

The Green Paper published based on this pre-accession assistance paved the political and administrative way to – and legitimate – dramatic structural changes in education sector, one of which is Decree No. 652 (ERI, 2013a). The Decree has been one of the milestones in the current ruling Justice and Development Party's (JDP) education agenda that was set out in the Urgent Action Plan in 2003, which promised strategic regulatory changes for quality control and assurance in public governance, as well as serious reduction in bureaucracy in line ministries. As a result, Monitoring and Evaluation Units under selected directorates at MoNE were established thanks to this decree. Although the legal commitment to better policies and programs has been welcomed (ERI, 2013a), the literature on the current role and utility of evaluation for Turkish educational decisions is extremely sparse. Little is known about the impact of EU's project on capacity building at MoNE, and in the absence of evidence, it is hard to

conclude that evaluation information foregrounds educational planning and programming in the country.

Notwithstanding the importance of these recent developments in Turkey, there is currently a lack of solid empirical base for unpacking what evaluation really offers to Turkey's educational decision making. Evidence from existing national and international literature is almost nonexistent. Available research on the Turkish education system, as far as M&E is concerned, has been limited to micro-level studies addressing student achievement scores at worst and curriculum evaluations at best (see Yüksel & Sağlam, 2012; Özdemir, 2009), with minimal attention paid to macro-level policy issues, normative questions of performance, quality, and development, as well as critical implications of global trend for good governance featuring a set of toolkit including M&E (Çelik, 2012). Although there is considerable anecdotal evidence about how some low and middle income countries (LMICs) – also known as the global South or the developing world – view, conceptualize, negotiate, materialize, and utilize evaluation for their governmental decisions, systematic studies of this phenomenon for Turkey are undocumented.

In stark contrast, during the last two decades, the evaluation community has witnessed a dramatic growth of the field in contexts outside of the global North (Chelimsky & Shadish, 1997). The notion that evaluation has an intrinsic value and is essential in any society has sparked the cross-national transfer of evaluation systems and practice (Fitzpatrick, Sanders, & Worthen, 2004). The assumption that evaluation advances human betterment ultimately generated a desire in the Northern-based donor

community to promote evaluation as a decision-making tool. This has led to efforts to build evaluation capacity in LMICs where a culture of evaluation is historically missing (Carden & Alkin, 2012; Schwandt et al., 2013; Porter, 2013). The momentum for promoting evaluation as a governing tool in the global South has reached a peak with the designation of 2015 as the International Year of Evaluation by EvalPartners – an international initiative to strengthen national evaluation capacities worldwide (http://mymande.org/evalyear). Although the empirical investigation of the impact of evaluation capacity building (ECB) in the developing world has been largely missing in the literature, existing anecdotal evidence suggests that the donor community has helped demystify evaluation practice in many LMICs (Mackay, 2009). Shifting the focus from the dominant Western model to indigenous systems, some scholars and practitioners probed the meaning and boundaries of evaluation concept and practice in developing countries (Carden & Alkin, 2012; Furubo, Rist, & Sandahl, 2002; Russon & Russon, 2000). Despite this bourgeoning literature, evaluation is considered an emerging construct in many LMICs, including Turkey, which requires further investigation into the value of evaluation in a developing country decision making context.

This sentiment was recently brought into full focus during a scholarly discussion on the future of evaluation in a forum of the *American Journal of Evaluation*, calling for a critical examination of what evaluation really offers to economically developing societies. "Without a clear value proposition," Leviton (2014), a former president of American Evaluation Association (AEA) argued, "evaluation will continue to be the orphan and the guest in organizations […] If organizations do not evaluate, we assume

they have limited evaluation capacity, but it may be that our value proposition simply does not impress them" (pp. 91-92). Leviton's (2014) question of whether evaluation's value for decision making is concise and compelling for individuals and organizations is apt. Evaluation is praised as a highly valued commodity for public policies and programs, but how its role and utility are viewed in different contextual – individual, organizational, national, regional, continental – settings remain obscure in the mainstream literature. Given Turkey's voluntary urge to modernize her system of governance in compliance with the global standards, meanings, margins, and negotiations of evaluation's conceptual and practical place in the Turkish education decision domain will be illuminative.

The Purpose of the Study and Research Questions

As the field of evaluation evolves globally and cuts across many geographies, histories, political systems, and most especially cultures, calls for the need to reflect on and be critical of our standards, assumptions, and values related to the concept and practice of evaluation have rightfully increased (Smith, 2012; SenGupta, Hopson, & Thompson-Robinson, 2004; Kirkhart, 2005; Hopson, 2001; Mertens, 2007). Empirical research is needed to increase the level of knowledge about the role and utility of evaluation in national decision making.

Michael Quinn Patton (2012), an evaluation pioneer both in the global Northern and Southern contexts, argued that different people attach different meanings to evaluation based on their perceptions, past experiences, and feelings related to the term. Recognizing the contextualized meanings and practice of evaluation, Patton (2012) argued that, "Research on *readiness for evaluation* has shown that valuing evaluation is a

necessary condition for evaluation use. Valuing evaluation cannot be taken for granted.

Nor does it happen naturally. Users' commitment to evaluation is typically fragile and

often whimsical…" (pp. 15-16, *italics in original*). As a result, Patton suggests assessing

perceptions, attitudes, and behaviors related to evaluation to investigate shared

understandings and purposes for evaluation in any given context. Following Patton's

lead, this study investigates the value of evaluation as a decision-making tool within the

Turkish education decision domain from the perspectives of key Turkish stakeholders –

government officials, academics, and civil society representatives – and official

statements by answering four central, interrelated questions:

1. To what extent and in what ways is evaluation utilized in the Turkish education
   decision domain?

2. How is evaluation conceptualized in the Turkish education decision domain?

3. What is the need and capacity for evaluation in the Turkish education decision
   domain?

4. What contextual factors facilitate or hinder the value of evaluation for Turkey's
   educational decision making?

To answer these questions, this study adopts two conventions from Weiss (1998)

concerning the definition of evaluation and the object of evaluation. First and foremost,

evaluation scholars have provided many definitions and discussed several purposes for

evaluation (e.g., King & Stevahn, 2012; Patton, 2012; Stufflebeam & Shinkfield, 2007;

Rossi, Freeman, & Lipsey, 2004). The most widely used definition of evaluation is "the

systematic process of determining the merit, worth, or value of something" (Scriven,

1991, p. 139). Many practitioners in LMICs – similar to this study – favor Carol Weiss'

definition of program/policy evaluation (see UNDP, 2011): "the systematic assessment of

the operation and/or outcomes of a program or policy, compared to a set of explicit or

implicit standards as a means of contributing to the improvement of the program or

policy" (Weiss, 1998a, p. 4). Here, it is vital to distinguish evaluation from other

research. Weiss (1998a) provided a comprehensive comparison of these two inquiry

traditions. For our purposes – evaluation for educational decision making – her insight

about the utility and client is more immediately relevant. Weiss (1998a) posits that

evaluations are conducted with *use* in mind for a specific *client* who has decisions to

make. Evaluations – unlike basic research – are intended to be used by policy or program

communities who need information to ground their decisions and interventions. By the

same token, this study purposefully separates evaluation from research by using the term

evaluation strictly within the context of government interventions that are targeted by the

very evaluation for a valued, purposeful use.

The second convention that this study adopts from Weiss (1998a) is to generally

talk about the evaluation of public (education) policies, which will occasionally

encompass national programs, projects, and their components. *Public policy* is simply a

government intervention that is intended to bring about a change based on specific

objectives (Dye, 2008). Education policy is a multidimensional concept encompassing all

endeavors of state actors, including programs, projects, decisions, discourses and

objectives, geared towards ameliorating a specific education problem or/and regulating

the lives of citizens (Keser-Aschenberger, 2012). *Program* can be defined in multiple

ways. The Joint Committee on Standards for Educational Evaluation defines program as "activities that are provided on a continuous basis" (Sanders, 1994, p. 3). Fitzpatrick, Worthen, and Sanders (2004) define program as "a complex of people, organization, management, and resources that collectively make up a continuing endeavor to reach some particular educational, social, or commercial goal" (p. 54). The authors submit that while some programs have identifiable boundaries, goals, administrators, budget, and procedures, some are multi-level, spanning across organizational and geographical boundaries, which make for evaluators harder to define. Given these definitions, to Weiss (1998a), the rationale for collapsing the categories of the evaluand (i.e., programs, policies, projects) – highly applicable for the purposes of the current study – is that "we can evaluate national programs, local projects, or subproject components, using the same basic methods" (p. 7). Additionally, due to the lack of literature and prior research on the role of evaluation in Turkey, the current study will cast a wider net to capture a range of issues and opportunities in policy, program, and project action settings.

Grounded on these conventions, the study utilized an exploratory single-case study with an embedded design to develop and refine theoretical propositions about the value of evaluation in a developing country decision-making context, drawing from the country case of Turkey. The case study approach served to provide rich and holistic exploration about the phenomenon in its real-life context by allowing for triangulation and corroboration among semi-structured interviews, document reviews, and reflexive memos. The aims of this study are heuristic, so it focused more on exploring relationships rather than on testing hypotheses.

Conceptual Framework

   This study is based on the premise that, like any other social endeavor,

evaluations are situated in cultural, political, social, and historical contexts that inform

and shape the design and implementation of policy or program alternatives that the very

evaluations seek to address (Patton, 2012). The American Evaluation Association (AEA)

(2009) defines context as follows:

> Context typically refers to the setting (time and place) and broader environment in which
> the focus of the evaluation is located. Context also can refer to the historical context of
> the problem or phenomenon that the program or policy targets as well as the policy and
> decision-making context enveloping the evaluation. Context has multiple layers and is
> dynamic, changing over time. (as cited in Patton, 2012, p. 36)

Evaluation scholars contend that the value of evaluation depends on context, and as such,

evaluations are value-laden (Chouinard & Cousins, 2009; SenGupta, Hopson, &

Thompson-Robinson, 2004; Mertens, 2008; Hood, Hopson, & Frierson, 2005; Patton,

1985). Indeed, "evaluation, itself, constitutes a culture, of sorts" (Patton, 2012, p. 50).

The issue of context in evaluation, as a result, problematizes the applicability of Western

cultural frameworks in non-Western settings (Mertens & Hopson, 2006). Evidence from

a wide variety of evaluation studies converges to suggest that the inquiry traditions of the

white, majority Western culture may compromise the interests of underrepresented

groups – low and middle income countries in this case – due to a widespread failure to

appreciate these groups' ontological and epistemological assumptions and cultural

nuances (Smith, 2012; Kirkhart, 2005; Kawakami, Aton, Cram, Lai, & Porima, 2008;

Bishop, 2012; Merryfield, 1985). To challenge the status quo in the field of evaluation

that tends to privilege modern, Western (European-North American) ontological

assumptions, and to increase the contextual credibility and relevance of evaluation

practice, many evaluation scholars advocate for the use of non-Eurocentric evaluation approaches that are grounded in a particular cultural, historic, and political milieu, and done by and for the community composing the very context (Pihama, Cram, & Walker, 2002; Hopson, Bledsoe, & Kirkhart, 2012; Mertens, 2008, 2009).

Despite the rampant concern for contextual differences, existing literature demonstrates a common growing interest in using evaluation as a decision-making tool for designing, implementing, and improving organizational goals at the local, national, and international levels (Fitzpatrick, Sanders, & Worthen, 2004). Western researchers have long argued that evaluations are influential forces to improve public services, programs, and policies (Segone, 2008; Mark, Henry, & Julnes, 2000; Scriven, 1991). They contend that evaluations contribute to institutional learning and effectiveness of decisions. Evaluation processes and findings might create changes in thinking about programs' and policies' design, implementation, logic, and desired outcomes, and ultimately generate shifts in action by building a solid knowledge base (Weiss, 1998a; King, 2008; Preskill, 2008; Patton, 2012). Fitzpatrick et al. (2004) summarized the significance of evaluations in decision making:

> Evaluation serves to identify strengths and weaknesses, highlight the good, and expose the faulty, but it cannot singlehandedly correct problems, for that is the role of management and other stakeholders, using evaluation findings as one tool that will help them in that process. (p. 27)

Given evaluation's significance, cross-national transfer of evaluation as a form of inquiry is taken as almost irresistible (Smith et al., 2011; Vidueira, Diez-Puente, & Afonso, 2013). Concerted efforts by many Northern and some Southern institutions and evaluators to build evaluation systems and practice in developing countries have

significantly contributed to this expansion. Numerous sessions, workshops, and conferences have been organized to build evaluation capacity in developing country governments, and many national evaluation organizations and associations have been established (Mertens & Russon, 2000). EvalPartners, an international evaluation partnership initiative to strengthen civil society evaluation capacities to influence public policy based on evidence, attempted to map existing Voluntary Organizations for Professional Evaluation (VOPEs) around the world and found some information on a total of 158 VOPEs, out of which 135 are at national level, while 23 are at regional and international levels (e-mail communication, Segone, January 2013). Some LMICs have established government-wide evaluation systems to improve their public programs and policies (e.g., Brazil, Korea, Mexico, etc.) (See UNDP, 2011). Most recently, an emerging interest in establishing evaluation as a profession in developing countries beyond development assistance has evolved (Carden, 2010). As a result, the field of evaluation in the 21st century has been characterized by its international and cross-cultural expansion (Patton, 2010).

The cross-national journey of evaluation as a form of inquiry combined with methods and systems, however, has ignited a critical debate among scholars within and beyond the field of evaluation. The seemingly global trend for borrowing and lending the evaluation concept and practice across countries coincides with a global trend in new public management, highlighting evidence-based practice and establishing effectiveness, efficiency, and accountability as common denominators (Clarke, 2008). This momentum for promoting good governance, currently regarded as key to successful economic and

social development in the global South, has pressured countries to undertake performance assessments and impact evaluations of national initiatives specifically sponsored by development aid. The pursuit for effective and sustainable development practice soon motivated donor agencies (multilateral or bilateral) and countries to fund evaluation capacity building activities, contributing to the evolution of evaluation systems and practice, hence evidence-based decision making in LMICs (Mackay, 2002; Picciotto, 2003). According to some critical researchers in the field of evaluation, however, the dominance of northern-based institutions' values and priorities might disable learning from evaluation for in-country decision making (Hay, 2010), and their warning has not been unwarranted. The trend for good governance is implicated with the neo-liberalization of development policy incentivized by such global actors as the IMF (International Monetary Fund), World Bank, OECD (the Organization for Economic Cooperation and Development), or UNDP (the United Nations Development Programme) in the aftermath of Post-Washington consensus (Steiner-Khamsi, 2004). A group of scholars in the field of comparative education problematized these global agencies' efforts to standardize, proceduralize, commodify, and market educational practices around the world through systems of global monitoring and evaluation that are believed to promote hegemonic standards of quality and quantity (Soudien, 2011; Valverde, 2014).

Alerted by the diffusion and internalization of Western epistemological imperatives into developing country context through evaluation practice that prioritizes a particular form of reality, an emerging group of evaluation scholars has pulled indigenous ways of knowing out of obscurity while discussing the practice of borrowing and lending

evaluation ideas and systems across nations. These scholars endeavor to untie the association between development and evaluation to dispel the hegemonic, standardizing and normalizing tendencies of everything global, and highlight the variability and complexity of context, which does not accommodate a linear, apolitical, and non-cultural procedure to specify, negotiate, design, implement, and improve policy alternatives (Hay, 2010; Smith, 2012). They pose and address the critical question of why, how, and under what circumstances is evaluation being promoted in the global South (Hopson, Kirkhart, & Bledsoe, 2012). Put differently, this bourgeoning literature draws attention to what is at stakes while evaluation systems trespass into unique territories. They advocate for the developing country people to take full ownership of their decision-making process and build bottom-up evaluation culture and capabilities for and by their own people in ways that are responsive to their contextual realities (Schwandt et al., 2013; Carden, 2007).

Turkey provides a good illustration of this debate. Geographically located in the Middle East and institutionally closer to the global North, Turkey provides a unique case for exploring the value of evaluation for governmental decision making. Turkey is often characterized as a bridge between the global North and South or East and West not only because of its geographical condition, but also its cultural and historical heritage (Davutoğlu, 2008; Gök, 2011). This western/eastern, northern/southern, "European/Asian geographical and cultural divide" embedded in the fabric of the society still continues to this date (Cooper, 2002, p. 589). The modern Turkish state emerged in 1923 "out of the ashes of the Ottoman Empire" in the aftermath of World War I (Ergil, 2000, p. 49), and Mustafa Kemal Atatürk (aka the father of all Turks) established the country based on

Western political ideals from the global North (Küçükcan, 2003). Turkey has been a member of OECD since 1961; associated with the European Economic Community (EEC) since 1963; a candidate for the EU membership since 2005; ranked among the 20 largest economies of the world in 2013 (CIA Factbook, 2013); and ranked at the 37th place in the 2013 World Competitiveness Scoreboard. Although the newly created Turkish state pronounced westernization and modernization as its foundational elements, the country has eclectically blended the ideological and cultural elements from the South into its social, political, and educational contours (Şişman, 2011).

Similar to other LMICs, Turkey's quest for development closely followed the global capital that in return inscribed particular structures for public governance, including evaluation systems. Official Development Assistance (ODA) has long contributed to Turkey's development and booming free-market economy – accompanied by a democratic, secular, republican regime – since the 1960s. Providers of ODA have included individual countries such as the U.S. and Germany, as well as bilateral and multilateral agencies such as the World Bank, UNDP, and the International Bank for Reconstruction and Development (IRDB) (Murphy & Sazak, 2012). In return, aid management has demanded that Turkey participate in rigorous M&E efforts to demonstrate effectiveness and efficiency of development assistance. As a part of the policy community, Northern donors invested their resources to have softened the process of adopting evaluation as a solution to governmental problems in Turkey largely through evaluations of aid projects. These efforts have informed and framed the conceptual and practical grounding of evaluation practice in the country (e.g., Stout, 2010).

Despite their ongoing efforts to build capacity, donors' success in installing evaluation mentality into Turkish governance has been limited (Şişman, 2011; Aslan, Küçüker, & Adıgüzel, 2012). Available evidence suggests that – borrowing Wadsworth's (2001) words – Turkish educational authorities have utilized donor evaluations to *check* and monitor (audit review) rather than to *seek* and improve programs and policies (formative inquiry) due to the country's lack of history with the concept and practice of M&E (UNDP, 2010; World Bank, 2012). As Turkey aims to play a bigger role in multilateral platforms, donor agencies pose the rise and fall of evaluation in Turkish governmental agenda as almost a threat to her competitive power (World Bank, 2012; European Commission, 2007, 2012; Murphy & Sazak, 2012). In the meantime, Turkish scholars and thought leaders continuously refocus attention on the role of Turkey's unique context in negotiating the terms of educational governance to produce responsive and culturally viable policies and programs without blindingly borrowing international concepts and practices (İnal & Akkaymak, 2012; Şişman, 2011).

Taking into account the distinct historical, political, and cultural context for educational decisions, policies, and programs in Turkey, this study argues that the value of evaluation for Turkey's educational decision making manifests unique features that pose and address challenging questions regarding the formations of the concept, practice, and field of evaluation in the global South. Thus, the seminal evaluation literature on context and culture (for instance Smith, 2012; LaFrance & Nichols, 2008; Greene, 2005; Conner, Fitzpatrick, & Rog, 2012; Hood, Hopson, & Frierson, 2005) provides the theoretical support for the current study to foreground contextual elements, including but

not limited to spatial location, historic structures, political and societal realities, and cultural traditions, in making meanings of evaluation in Turkish imaginary.

Significance of the Study

This research has important implications for education practices in Turkey, as well as implications for the field of evaluation. First and foremost, this study seeks to contribute to educational policy making in Turkey. Despite many education reforms Turkey has passed over the past few decades, scholars argue that educational policies and programs have fallen short of remedying educational problems (Çelik & Gür, 2013). Certainly there are numerous reasons and determinants of why the education reforms and program are not working, but one challenge that prevents Turkey from effectively addressing educational problems is the gap in the base of knowledge about which programs work best to improve educational programs for whom and under what circumstances (Gür, 2011). This is a significant problem because decisions based on inadequate information about programs' merit may lead to poor use of social resources (Weiss, 1998a). Thus, it is worthwhile to provide decision makers with systematic information as to whether the various educational programs are worth the money they cost, whether they should be continued, and how they can be improved to meet the societal needs (Bamberger, Rugh, & Mabry, 2012). By exploring the utility of evaluation, Turkish policymakers may better gear state policies and reforms towards a more responsive culture of decision making.

In addition, this study seeks to expand the knowledge base surrounding the efforts to build evaluation systems and practice in low and middle-income countries to improve

national decision making. Thus far, only a few studies have addressed the value of evaluation systems and practice in decision-making contexts from the developing country perspective. As noted by Hay (2010): "Evaluation research, innovation, and leadership should not remain exclusive to northern based institutions. We need to examine how evaluation research is developing and the role southern evaluators and organizations are playing in this process" (p. 226). Thus, without a clear understanding of how a developing country views the value of evaluation as a decision-making tool to improve its educational practices, the field lacks future directions about how to contribute to social betterment worldwide. Located between the global North and the global South whose evaluation history, needs, and activities have been relatively distinct, Turkey offers a challenging case to investigate this phenomenon. Borrowing Hay's (2010) terms, Turkey may be the testing ground for illuminating future directions of the field in the developing country context. Thus, this research will extend the knowledge base about the nature and parameters of evaluation conceptualization and utilization in different country settings and broaden the scope taken by Western researchers and practitioners.

Furthermore, this research answers calls for studying the concept of evaluation outside of the Global North (Carden & Alkin, 2012; Hopson, 2001). Fitzpatrick et al. (2004) believe that in the twenty-first century, Western evaluators will increasingly hear and learn from their counterparts' practices in other countries that will only work to strengthen their own practice. Thus, a deeper understanding into the concept of evaluation in Turkey will not only facilitate the solutions for policy problems in the global South, but may also motivate Northern evaluators to unlearn and relearn some of

23

their assumptions, methods, and approaches to evaluation (Mertens, 2009). The proposed study breaks new ground by assembling the global North and South-based evaluation literature, which embraces, in some respects, distinct models and approaches.

Delimitations

There are number of limitations and caveats to the proposed rationale and conceptual design of this study. First and foremost, the scant literature on the concept and practice of evaluation produced by and for the low and middle income country stakeholders poses a limitation. This research borrows literature from both Western and non-Western scholars (where applicable) to probe the boundaries of the concept and practice of evaluation in distinct country contexts. However, relatively limited, local evaluation practice and literature in the global South may elevate the voices of Western notions of evaluation practice in the current study. While sharing his lived experiences with efforts to develop *African Evaluation Standards*, Hopson (2001) touched on this limitation, and invited evaluation practitioners to challenge the Western ethnocentrism in our field. He noted:

> From my point of view, the main rationale for the development of African evaluation standards relates to redefining foundations of the field as they relate to the unique contributions to be made by culturally and ethnically diverse communities- whether globally or locally- and recreating methodologies and paradigms that have historically served to cripple or debilitate communities that have been underserved or marginalized. (p. 376)

Following Hopson's lead, my purpose is not to impose a Western concept (i.e., monitoring and evaluation) on Turkish stakeholders, but rather to try to understand how they approach evaluation, which is praised as an influential decision-making tool in

several country contexts, and understand what unique perspectives they offer to inform and shape the foundations of the field in the twenty-first century.

Second, although the study greatly benefits from a distinction between the global North and South in its effort to intentionally refocus the reader's attention on context, the tendency to essentialize these terms and places remains as a challenge. By positioning Turkey in between these – mainly – theoretical spaces, the attention is directed towards the fluidity and non-fixity of these constructs so that the findings can motivate the reader to reconsider his/her assumptions and expectations related to dominant narratives.

Third, the present study focuses specifically on the educational decision making context. Hence, the results of this study may not be fully applicable to other decision domains in Turkey (i.e., health, transportation, employment, development, etc.). However, this study is the first empirical account of the value of evaluation for public policymaking in the country. Thus, it is possible that the findings may foreground future research by highlighting critical questions for the entirety of governmental decision making.

Parallel to this delimitation, the reader should be reminded of a serious caveat inherently posed by single case studies; that is, the case characteristics may change during or after the course of the study that may weaken some of the basic assumptions or even major findings. With this caveat in mind, and following the guidelines for a single-case study, the current research does not adhere to statistical generalization whereby a sample can help us make inferences about the larger population (Yin, 1994). The criteria to judge the success of exploring evaluation's value in Turkey is the illuminative power

and critical outlook of the findings at this point in time, without generalizing to other cases but rather a broader theory about the value of evaluation. Thus, the potential changes in the case (i.e., Turkey's education decision domain) will only provide more room for a critical research without delegitimizing the findings of this study.

Organization of the Dissertation

This dissertation is divided into six chapters. Chapter 1 has described the problem, significance of this problem, research purposes, and questions. Chapter 2 presents a literature review on the subject of the role and utility of evaluation as a decision-making tool in low and middle-income countries at large. The analysis draws on evaluation literature from both the global North and South. Chapter 3 outlines the research methodology and rationale for a case study design. Chapter 4 lays the groundwork for the educational decision making context in Turkey, introducing actors, processes, and motivators of current educational governance and reform agenda in the country. This chapter also illuminates the role of evaluation in the education decision domain as reported in major official documents. Chapter 5 details research findings derived from the analysis of the data, drawing from individual interviews, document review, and reflexive memos. Finally, Chapter 6 summarizes the major study conclusions for the value of evaluation for Turkish educational decision making. The chapter also offers implications for evaluation scholars and practitioners around the world regarding the future directions for research and practice in building national evaluation systems and expanding the field worldwide.

Chapter II: Literature Review

Introduction

      This chapter explicates evaluation studies and international development literature that provide the foundation for the conceptual framework presented in the first chapter. The literature on the value of evaluation as a decision-making tool in low and middle income countries (LMICs), including Turkey, is scarce. Thus, this review benefits mostly from publications in the databases of aid organizations such as the World Bank, UNICEF, and OECD, as well as the Northern-American and European based journals such as the *American Journal of Evaluation*, *Evaluation*, and *Canadian Journal of Program Evaluation*, *New Directions for Evaluation*, and *Evaluation and Program Planning*. Overall, the review of literature uncovered four focal areas to understand the many aspects of the value of evaluation: context for evaluation, use of evaluation for decision making, evaluation capacity building in the developing world, and indigenous evaluation practice.

      In keeping line with the review of literature, this research is positioned within the intersection of critical and constructivist paradigms. Critical theory assumes that a reality is produced and mediated by power relations in politically, culturally, socially and historically situated contexts (Guba & Lincoln, 1998; Kincheloe & McLaren, 2011). Thus, the current research will borrow from the evaluation literature produced and circulated by the Northern-based, aid organizations to uncover the reproduction of political discourse about the value of evaluation that may be routinized by established structures and are taken as real in the global South. The constructivist paradigm assumes that there is no single truth about the value of evaluation: realities are indeed multiple, socially constructed, and specific to context and individuals who attach specific meanings

to evaluation activities (Guba & Lincoln, 1998). Therefore, the present paper intends to synthesize the global North- and South-based literature in a circumscribed manner and use their key ideas and concepts to uncover conflicting interests and associated realities of different country people related to evaluation, and eventually to distill a more sophisticated consensus construction of program evaluation.

This chapter will first focus on the significance of context in affecting the meaning, the role, and the utility of evaluation, hence its value as a decision-making tool in non-Northern country contexts. Then the review will describe the purposes, uses, and significance of evaluation in decision-making contexts at large. This section will draw largely on empirical and theoretical research studies conducted in Western organizational settings. Two broad bodies of literature will follow. One is related to Northern-based and created aid organizations' evaluation capacity building efforts in the developing world. The second body of literature concerns indigenous, bottom-up evaluation field building efforts in LMICs. I will describe definitions and discuss main themes, strengths and weaknesses related to each body of literature. The last section will conclude with a synthesis of the main points in the reviewed literature.

Context Matters

This chapter is premised on the understanding that context matters in defining the role and value of evaluation in a society. Evaluation scholars have long argued that political, social, historical and cultural context informs, shapes, and influences evaluation practice (e.g., the selection of questions, appropriateness of methods, dissemination of results, and uses of findings) (Conner, Fitzpatrick, & Rog, 2012; Hopson, Kirkhart, & Bledsoe, 2011; Alkin, 2004a; Chouinard & Cousins, 2009; SenGupta, Hopson, &

Thompson-Robinson, 2004; King & Volkov, 2005). As noted by Grudens-Schuck (2003),

"It is important that we consider that evaluation, like other events and behaviors, is part

of culture and has been developed to have meanings particular to time and place" (p. 24).

Established scholars of the field have recognized that evaluation is a social intervention;

hence contend that evaluation reality is produced in politically, culturally, socially and

historically situated contexts (Guba & Lincoln, 1989; Greene, 2006; Mertens, 2008;

LaFrance & Nichols, 2008). Truth about the value and utility of evaluation can never be

isolated from a domain of political discourses, cultural values and historical relations

(Bamberger, 1991; Smith, 2012; Hood, Hopson, & Frierson, 2005). As a result, they

argue that context plays an essential role in grounding and validating the concept of

evaluation in a particular setting for a particular group of people, as well as the ways in

which it can be conducted and used (Conner, Fitzpatrick, & Rog, 2012).

Greene (2005) provided a widely recognized definition of context: "Broadly

speaking, context refers to the setting within which the evaluand (the program, policy, or

product being evaluated) and thus the evaluation are situated. Context is the site, location,

environment, or milieu for a given evaluand" (p. 85). Both the *Program Evaluation*

Standards (Joint Committee on Standards for Educational Evaluation, 2011), and

Guiding Principles for Evaluators (American Evaluation Association [AEA], 2004)

highlight the importance of context in evaluation practice. While some evaluation

scholars focus more on the decision-making context (e.g., Stufflebeam, 1968; Weiss,

1972, 1973; Stake, 1983; Alkin, 1969; Patton, 2012), others have focused on the cultural

context (e.g., Mertens & Hopson, 2006; Kirkhart, 2005, 2010; Kawakami, Aton, Cram,

Lai, & Porima, 2008; Chouinard & Cousins, 2009; LaFrance, Nichols, & Kirkhart, 2012).

Regardless of the focus, all these scholars encouraged evaluation practitioners to *carefully* examine and be responsive to cultural norms and values; national sensitivities, political systems, and interests; the styles of decision making and communication; local knowledge, and social structures. Many evaluation scholars have illuminated how a *purposeful* understanding of and *explicit* attention to the context of evaluation can improve the relevance, utility, quality and rigor of evaluations (Rog, 2012; Greene, 2005; LaFrance, Nichols, & Kirkhart, 2012).

It is now widely accepted in the field of evaluation that context of evaluation is multi-layered and complex (Greene, 2005), and scholars have endeavored to develop descriptive models and approaches to capture the most salient dimensions. Rog (2012) provides one of the most recent models. The author describes five areas of context to consider while conducting context-sensitive as opposed to methods-first evaluations: the nature and circumstances of the problem, the broader decision-making context, the nature and dynamics of the intervention, the framework for the evaluation itself, and the broader environmental factors. Within each of these areas, Rog calls out specific dimensions to attend to: the physical, organizational, political, historical, social, traditional, and cultural. Conner, Fitzpatrick, and Rog (2012) proposed a three-step (planning, implementation and use) approach to context assessment (CA) that incorporates Rog's (2012) five elements. They present key evaluation questions and issues to be addressed under each element within each step. An eclectic combination of their key questions especially from the areas of broad environment and decision making can be fruitful to understand how the role and value of evaluation will justifiably vary depending on the context (see Table 1).

Table 1. Context Assessment

Area	Mainstreaming Evaluation
Broad environment around evaluation	• What are the important historical, social, and cultural elements of the country to which the evaluation is introduced? • Are there political and social views that affect perspectives on evaluation?
Broad decision-making arena	• Who are the main decision makers? • What are their views, values, and history about evaluation? • How do their values, position, or history affect their use of information (or evaluation)? • What is the larger political culture in which they work? • What are different expectations of their job? • What are different expectations of citizens regarding their job? • What are the political expectations for evaluation?

Note. Adapted from Conner, Fitzpatrick, and Rog (2012, pp. 96-102)

In their quest to bring context from background to foreground, questions such as these made evaluation practitioners increasingly aware of their assumptions related to the concept and practice of evaluation in different country contexts (Furubo, Rist, & Sandahl, 2002; Nevo, 1982; Merryfield, 1985). Indeed, some scholars have made a call to engage with people from different countries and cultures to probe the meaning of evaluation and how evaluation is compatible with their particular context (Hay, 2010; Carden & Alkin, 2012; Williams & Hawkes, 2003; Hopson, 2001). To date, however, few studies have empirically investigated different conceptualization and interpretations of evaluation in different county contexts.

The International Atlas of Evaluation by Furubo, Rist, and Sandahl (2002) may be the best-known, comprehensive, one-stop document of evaluation cultures (or lack thereof) in more than twenty-one countries around the world spanning five continents, including such LMICs – also known as the global South or the developing world – as China, Korea, and Zimbabwe. The authors describe these countries' evaluation systems,

how evaluation *took off* and contextual forces (political, socio-economic, cultural, administrative, the role of the state, civil society, etc.) that affect the development and consequences of such systems. The researchers also debate about the meaning of evaluation in different national contexts. They acknowledge that the definition of evaluation does not stay the same and note:

> It is better for each author to tell the reader what he or she considers to be evaluation or not. The purpose of this book is […] to let each author describe and analyze current developments in what in the context of the country or institution is deemed to be evaluation, not to establish a definition that is more correct than any other. (p. 4)

Each chapter of this book provides new – and mostly the first – insights into how evaluation is understood, conducted, used and institutionalized differently in each selected country.

Parallel to this, Carden and Alkin (2012) provide one of the most recent articulations of evaluation practice in LMICs. The authors describe three methodologies (adopted, adapted and indigenous) used in LMICs to facilitate the location of their evaluation practices on the theory tree described in Evaluation Roots by Alkin (2004b, 2012). First, the authors recognize that evaluation practice was first expanded to LMICs through aid organizations as a means to deliver their services, hence the expansion of evaluation practice to LMICs was originally a "technology transfer" (p. 106) of adopted, context-neutral methodologies such as Logical Framework Analysis, results-based management, evidence-based practice, and impact evaluations with randomized control trials, which "for better or worse, does not see a need for adaptation to local context" (p. 108). Second, to Carden and Alkin (2012), this "colonial role of evaluation" (p. 108) facilitated by the donor community has created a need in LMICs to contextualize the methodologies and address local needs. Some of the evaluation approaches that LMICs

adapt were participatory approaches, Patton's (2010) Developmental Evaluation, and Realist Evaluation by Pawson and Tilley (1997), to name a few. Although these methodologies have still been developed in the global North, the authors contend that their uses in LMICs are contextualized. Finally, some indigenous methodologies were developed and evolved in the global South as the byproduct of a national dialogue that reframes evaluation practice grounded in context and culture such as systematization (Sistematización) in Latin America. Grounding their findings on a review of relevant literature and professional experience, this scholarly piece by Carden and Alkin (2012) drew attention to a gap of empirical knowledge in the field of evaluation regarding how the concept and practice of evaluation is transferred, negotiated, and applied in different developing country setting.

Accompanying the discussions regarding the differentiated meanings and applications of evaluation in different country contexts has been a more general debate about the applicability of Western (evaluation) models and frameworks in non-Western settings (Mertens & Hopson, 2006; Smith, 2012; Bishop, 2012). Many Western evaluation scholars and practitioners have recognized that evaluation practice was first expanded to LMICs through Northern-based aid organizations as a means to deliver their services (Carden & Alkin, 2012; Hay, 2010; Furubo, Rist, & Sandahl, 2002; Khan, 1998). Thus, aid organizations' efforts to disseminate evaluation systems and practice in developing countries have implications for the ascribed value of evaluation for decision-making processes. Evidence from a wide variety of evaluation studies converges to argue that many developing countries consider evaluation a donor-driven activity without any value to their specific learning and information needs (Khan, 1998; Ba Tall, 2011). This

imposed use of evaluations arguably reduced the opportunities to increase and sustain

national evaluation capacity to address national information needs and improve decisions.

As noted by Hay (2010), Northern-based and created aid organizations' dominance over

the field of evaluation "created and reinforced inequalities in the global evaluation field

by overemphasizing the values, perspectives, and priorities from the North and

underemphasizing those from the South" (p. 224).

Despite the widely held concerns about the inapplicability of Western evaluation

frameworks in the global South, the global expansion of evaluation practice has revived a

significant debate about the role and utility of evaluation as a decision making tool in

LMICs. To provide plausible directions to the interplay between context and evaluation,

this study examines the value of evaluation in the developing country context by using

Turkey as a testing ground.

Evaluation for Decision Making

According to some evaluation scholars and researchers, the concept of evaluation

is inextricably linked to decision making (Chelimsky, 2006; Weiss, 1998a). This

utilitarian view of evaluation assumes that evaluations can provide useful information for

decision makers that help guide program or policy design, implementation, and

improvement, which is why evaluations can be appealing to managers, administrators, or

even policymakers (Fitzpatrick, Sanders, & Worthen, 2004). Stufflebeam (2004) is one of

the founders of decision-oriented approach to evaluation. The scholar developed the CIPP

(context, input, process and product) model for evaluators to collect four different types

of information for different decisions (Stufflebeam & Shinkfield, 2007). Similarly,

Alkin's (1969) UCLA Evaluation Model is also a common example of a decision-

oriented evaluation approach with its emphasis on providing useful information for decision makers. Patton's Utilization-Focused Evaluation (UFE) can also be considered under this utilitarian approach to evaluation because of Patton's interest in serving decision makers' information needs in different programmatic and organizational contexts (Patton, 2012). Despite the prevalence of decisions – its setting, structures, actors – in evaluation discussions, the decision-oriented approach to evaluation attracted criticism because of its tacit assumption that decision-making process is rational and linear (Fitzpatrick, Sanders, & Worthen, 2004; Patton, 2012). Accompanying this criticism is the more general debate about the disparate perceptions of purposes and uses of evaluation in decision-making processes.

Conceptions of evaluation. The field houses a plethora of conceptions of evaluation each of which prioritizes different philosophical and methodological assumptions and, sometimes, different purposes, and uses. Three decades ago, Nevo (1983) attempted to conceptualize evaluation for educational settings. Expanding on the literature that classifies evaluation approaches and models, the author defined ten major dimensions of evaluation: (1) the definition of evaluation, (2) its functions, (3) its objects, (4) information to be collected, (5) criteria to judge evaluation, (6) its audiences, (7) its process, (8) its methods, (9) characteristics of evaluators, and (10) standards to judge evaluation. Evaluation scholars and practitioners have engaged in heated debates about disparate perceptions of these dimensions such as the paradigm wars of the 1970s (Donaldson, 2008; Martin, 2005). Interest in considering evaluation in different ways, as a result, implies that different people assign different roles to evaluation and attach different values to the concept and practice (Fitzpatrick, Sanders, & Worthen, 2004).

Among Nevo's (1983) dimensions, the functions of evaluation always constitute a central piece in evaluation debates, which can be possibly if not exclusively characterized by two somewhat competing views: evaluation for accountability and evaluation for learning (Torres & Preskill, 2001; King, 2008; Chelimsky, 2006). The dichotomy between accountability (judgment) and learning (use) functions of evaluation is false according to King (2008) who cautions evaluators not to "split into two competing camps" (p. 154). While Chelimsky (2006) also criticizes a rigid adherence to one purpose of evaluation, the evaluation literature is replete with studies that promote the supremacy of one approach to evaluation over another (Preskill, 2008; Mertens, 2008; Greene, 2006). For example, Scriven (1967) repeatedly reported his conceptualization: evaluation has only one goal (making a judgment about the value of an object) and many roles (mainly, formative and summative). Scriven demanded evaluators to be objective and independent while making rational judgments about the value of an object because society at large (specifically consumers of education products) may waste their time and money for bad products whose values are not properly determined (Fitzpatrick, Sanders, & Worthen, 2004). Elsewhere, Scriven (1967) noted:

> Failure to make this rather obvious distinction between the roles and goals of evaluation is one of the factors that has led to the dilution of what is called evaluation to the point where it can no longer answer the questions which are its principal goal, questions about real merit or worth. (p. 62)

Some evaluation researchers rejected this *judgmental* purpose, arguing that Scriven's distinction highlights a negative aspect about evaluation (the pursuit of accountability) instead of its constructive features (learning and improvement) (Preskill & Torres, 2000; Fleishner & Christie, 2009; Mark, Henry, & Julnes, 2000; Weiss, 1998b). Elsewhere, many scholars emphasized several other valued purposes for

evaluations such as informed decision making, organizational learning, program

improvement, program development, advocacy, empowerment, social justice and

democracy (e.g., Preskill, 2008; King, 2008; Patton, 1998; Stufflebeam, 2004; Alkin &

Taut, 2003; Weiss, 1998b; Mertens, 2007; House & Howe, 2003; Greene, 2006; Hood &

Hopson, 2008; Chelimsky, 2006).

Recently, learning and improvement functions of evaluation have gained

prominence in the literature. Preskill (2008) defines evaluative learning as follows:

> The processes and findings of evaluations may result in thinking about a
> program's logic or theory of change; changes in a program's design and
> implementation; changes in program policy; changes in perception about the
> merit, worth, or significance of a program; or changes in attitudes concerning
> evaluation's potential value. (p. 129)

These researchers consider evaluation as a meaning-making process whereby different

stakeholders share, discuss, and contest their expectations, assumptions, beliefs, and

needs to create collective knowledge and improve their organizational practices (Preskill

& Torres, 2000). They highlight formative inquiry and the learning function of

evaluations for continuous improvement whereby evaluative thinking and evaluative

evidence (not only evaluation findings but also processes) become a part of the given

decision-making context to improve organizational practices over time (Preskill, 2008).

Supporting the learning function of evaluations, Patton (2001) is concerned that when

evaluations are mandated, then it will likely "yield surface adoption of evaluation and

mediocre work" (as cited in Williams & Hawkes, 2003, p. 65). He goes on to argue that

this will only serve accountability purposes at the expense of formative uses that are most

useful. By highlighting the distinction between the learning and accountability functions

of evaluation, these researchers draw attention to the particularities of the utility of

evaluation for decision making in a given context.

In sum, Patton (2012) summarizes these distinct purposes into six categorizes (see

Table 2), and argues that evaluation's purpose and use are defined by the intended users'

information needs, and priorities. Underlying his argument is the assumption that

contextual sensitivity is the key to understand the valued purpose of evaluation.

Table 2. Primary Purposes of Evaluation

Purpose of Evaluation	Focus	Evaluation Approaches	Primary Users
Summative, Judgment	To determine the overall value of the program	Summative evaluation, Impact Evaluation	Funders and policy-makers
Formative improvement, Learning	To improve the program	Formative evaluation, capacity-building	Program administrators and staff
Accountability	To demonstrate the efficient use of resources	Performance management, accreditation	Executive, legislative authority
Monitoring	To provide data for program management	Performance indicators, quality control	Program managers
Developmental	To adapt the program in complex environments	Real-time evaluation, developmental evaluation	Social innovators
Knowledge Generation	To identify patterns of effectiveness	Meta-analysis, lessons learned	Program designers

Note. Adapted from Patton (2012, pp. 129-132).

Uses of evaluation for decision making. Underlying the long-standing debate

about the functions of evaluation (accountability and learning) have been different

perceptions of the uses of evaluation for decision making, presumably the most

investigated topic in the evaluation literature (Mark & Henry, 2004; Christie, 2007).

Evaluation use and utilization are used interchangeably in the literature that refers to "the way in which an evaluation and information from evaluation impacts the program that is being evaluated" (Alkin & Taut, 2003). Although the researchers use different terminology at times (Almeida & Bascolo, 2006), the agreed-upon, prevalent uses of evaluation noted in the literature are three-tiered (Johnson et al., 2009).

First, *instrumental use* refers to using evaluation findings for immediate decision making to modify, expand, or terminate the evaluation object (product, program, policy or personnel) (Mark & Henry, 2004; Johnson et al., 2009). This use assumes a rational decision-making process where policymakers have access and desire for scientific evidence to use for their decision (Almeida & Bascolo, 2006). Second, *conceptual use* refers to indirect use of evaluation findings that illuminate policy problems and solutions in a new way and change our understanding. Carol Weiss (1998b) calls this *enlightenment* whereby the evaluation findings build knowledge and become a part of the policy discourse over time. Third, *symbolic use* refers to using evaluation findings to justify the existing practices, persuade others about certain positions or delay action in political arenas (Almeida & Bascolo, 2006; Weiss, 1979). Patton (1998) added *process use* to this list, arguing that mere participation in evaluation processes can also create learning to improve programs. Elsewhere, Weiss, Graham, and Birkeland (2005) drew attention to *imposed use* whereby the funder (i.e., government) requires use of scientific evidence for making programmatic decisions due to accountability concerns. Kirkhart (2000) elevated the debate to the next level and argued that use is a limited term to cover the potential effects of evaluation findings and process, hence proposed the term

influence instead, which is defined as "the capacity or power of persons or things to produce effects on others by intangible or direct means" (p.7).

Among these, the propositions of instrumental use are regarded the most contentious in the field of evaluation. Many evaluation scholars and practitioners have questioned the instrumental use of evaluation findings and processes for decision making, claiming that there are many sources of evidence available to decision makers; thus, it is unrealistic for evaluations to purport direct influence on decisions (Weiss, 1998a; Chelimsky, 2006). These scholars mainly question the direct use of scientific, sociological knowledge in the political arena, arguing that knowledge utilization does not take a form of immediate, direct application, but rather a longer and indirect transformation through various mechanisms (Balthasar & Rieder, 2000). The model created by Mark and Henry (2004) provides a picture of this complex mechanism (see Figure 1). Based on empirical investigations in social and behavioral sciences, the authors created a two-dimensional framework to illustrate the pathways (i.e., general influence, cognitive/affective, motivational, and behavioral) through which evaluations exercise influence at different levels of analysis (i.e., individual, interpersonal and collective). In line with Weiss' (1998) reservations about the instrumental use of evaluations, the scholars' model demonstrates that direct, immediate use of evaluation findings and processes may be more difficult than prescribed in the literature.
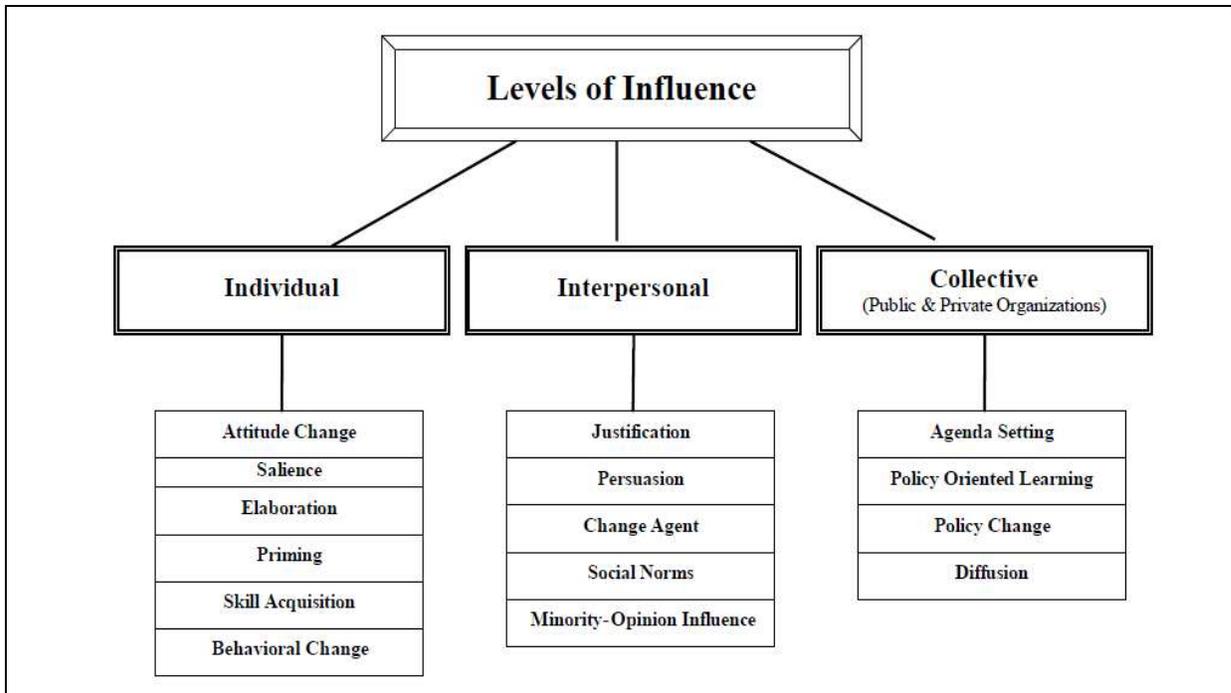
Figure 1. Mechanisms through which evaluation produces influence (Henry & Mark, 2003, p. 298).

The debate about the instrumental use of evaluation information is indeed linked to a larger discussion regarding evidence-based decision making. Changes in public administration culture in Western democracies over the last century have highlighted effectiveness and efficiency as common denominators in providing public services (Clarke, 2008). This shift triggered wide adoptions of results-based management and accountability mechanisms to demonstrate value for money and ultimately stipulated rigorous, scientific evaluations to produce credible, empirical evidence to be used for governmental decision making (Donaldson, 2008). Proponents of evidence-based policy and practice assumed that scientific evaluation evidence – presumably obtained from randomized control trials or quasi-experimental designs – will have a direct impact on the design and implementation of programs and policies, contributing to their betterment (Shadish, Cook, & Campbell, 2002).

Here, a brief discussion of policymaking theories is necessary to better understand the potential limitations of the use of evaluation evidence directly for governmental decision making. A quick scan of literature reveals at least two broad theoretical camps: rationalistic approach and incremental approach. According to rationalistic approach, policymaking follows a set of evolutionary steps or stages, and one of these steps is to measure the impact of the policy (Fowler, 2009; Simon, 2007; Hoy & Miskel, 2008). Similar stages of a policy cycle have been formulated by the proponents: agenda setting, policy formulation, policy implementation, policy evaluation, and policy change (Fowler, 2009; Theodoulou & Kofinis, 2004). While this approach essentializes evaluation as a necessary step in a policy cycle, it is widely critiqued for falling short of explaining whether and how policymakers are capable of accessing and assessing information regarding all existing alternative courses of action to identify the most optimal course of action. It is also critiqued for arbitrarily dividing the policy cycle into steps while underestimating the nonlinear interactions and feedback among the stages (Birkland, 2005; Heck, 2008). Thus, rational approach is believed to overestimate the possibility of gaining consensus on policy issues including problem definition and alternative specification based on available information, which might be provided by evaluation.

Incrementalism, on the other hand, is believed to embrace a less rigid and deterministic approach to policymaking, arguing that policies are not formulated once and for all towards a set-in stone objective (Anderson, 2006). Instead, policies are corrected along the way incrementally testing previous assumptions and reassessing objectives in moving forward. As opposed to rationality approach's rigid understanding of consensus and agreement based on available evidence and information, incrementalists

believe that their approach is much more realistic in obtaining agreement among various parties because of the likelihood of modifications later on. Without essentializing evaluation in policymaking, it is this later approach that provides more realistic insights about decision makers' likelihood of using evidence directly for their actions.

Parallel to this, many evaluation scholars also demonstrated the limitations of direct use of evaluation evidence for decision making (governmental or organizational) mostly due the political context. For example, Patton (2012) argues that this understanding "is based on the paradigm of logic and rationality that is inconsistent with how people take in and use data" (p. 11). Politics of evaluation have been a significant motivation for Patton to develop Utilization-Focused Evaluation (UFE) as he recognizes that politics in most cases trumps evaluation evidence in decision-making contexts. Another prominent critic of rational, evidence-based decision making, as well as instrumental use of evaluation information is Carol Weiss. Weiss (1998b) submits that evaluations may not have a direct effect on policy decisions due to various reasons. First and foremost, Weiss (1998b) posits that evaluation evidence competes against many other sources of information available to policy makers. Evaluation is not the primary source of evidence in policy arena (Weiss & Bucuvalas, 1980). The following poem summarizes the realist approach of Weiss and her colleagues (Weiss, Graham, & Birkeland, 2005) to use of evaluations for decision making:

Evaluation is fallible
Evaluation is but one source of evidence
Evidence is but one input into policy
Policy is but one influence on practice
Practice is but one influence on outcomes. (pp. 12-13)

Second and more importantly, Weiss argues that there is no single decision maker who easily welcomes evaluative evidence, and makes decisions without any conflict. To Weiss (1987), the policy decision-making context is indeed unstable, involving many decision makers who have different opinions, conflicting interests, and opposing needs that make it harder for evaluative evidence to directly inform decisions.

Many empirical studies illustrate Weiss' point. For example, Fleishner and Christie (2009) found supporting argument for Weiss' point from a cross-sectional survey of 1,140 U.S. American Evaluation Association members. The great majority of the sample agreed that decision makers do not always make their decisions based on evaluation evidence and that stakeholders reject evaluation conclusions based on their beliefs and values (Johnson et al., 2009). Similarly, Christie (2007) conducted a simulation study with 131 participants to understand how three different types of evaluation information (large scale evaluation study, case study data, and anecdotes) influence participants' immediate decisions (i.e., instrumental use). The author found that decision makers' information needs change depending on the political and social context within which they operate; this ultimately affects if and how they are willing to use evaluation information. Cook (1997) also explicated this point many years ago:

> The politician's prime goal is to be reelected rather than to respect technical evidence; that personal and party political ideology often entail that evidence is used in markedly selective ways; and that politicians experience a greater need to be a part of budget allocation rather than of program review. (pp. 40-41)

These studies point to a prevalent argument in the evaluation literature; that is, evaluations take place in political contexts (Weiss, 1998b; Mark, Henry, & Julnes, 2000; Greene, 2006; Datta, 2011). Weiss (1987) submits that programs that evaluations are concerned about are the very byproducts of politics, and evaluations make political

statements about the value of these programs and determine their fate. A review of knowledge utilization literature by Almeida and Bascolo (2006) suggested that there are several obstacles to using evidence in policy making: 1) lack of political will, 2) miscommunication or lack of interaction between researchers and policy makers, 3) lack of consensus in the research community regarding the solutions of certain policy problems, 4) timing of availability of research evidence, and 5) lack of means to make evidence available (media controls the communications channels). Chelimsky (2006) supported these assertions and argued that evaluators' ability to influence policy agenda depends on what they know and how they understand the policy environment within which evaluations and policy decisions are embedded.

Further, Young (2005) explored how political context can complicate knowledge utilization in many developing countries. It is worth noting that knowledge utilization literature is conceptually different from evaluation utilization because it emphasizes the use of social science research instead of findings from the implementation of social programs. Yet, Young (2005) touches on a common thread between the two, which is the use of credible, empirical findings. The author submits that the policy process in developing countries is chaotic and unpredictable, which makes it much harder for research and evaluation results to influence policy agenda (Young, 2005). Elsewhere, Court and Young (2003) developed a three-dimensional framework to illustrate factors that affect the use of evidence in international policy settings. Political context (political culture, civil and political rights and freedom, government officials' attitudes and perceptions, political pressures and institutional framework), the credibility and quality of evidence (timeliness and relevance of information for the policy problems), and links

(interaction and communication between government officials and researchers) determine the faith of research uptake in policy design and implementation. The authors believe that context is the key to determine the level of evaluation influence on policy change.

Contandriopoulos and Brousselle (2012) provided one of the most recent studies on the interplay between political context and evaluation. Based on a systematic review of knowledge exchange studies at policy-making and organizational levels, the authors developed a two-dimensional model to understand evaluation influence at the collective level. One dimension is called "issue polarization," whereby several users may have divergent opinions and competing interests about a given issue and access to several sources of information and knowledge; thus, they pick evidence that best suits their interests. As the issue polarization increases, rational decision making that uses scientific evidence for policy formulation or debates diminishes. The other dimension is called "cost-sharing equilibrium," which refers to who will bear the costs of information flow or/and knowledge exchange: users (who utilize evaluation evidence for their decision and can intervene in the knowledge exchange because of the formal positions they occupy) or producers (who produce the evaluation evidence). Similar to Mark and Henry (2004) and Weiss (1977), the authors argue that evaluation is not the only source of evidence or information for collective systems to use; thus, scientific information does not often directly affect collective systems (policy making or organizational behaviors).

Recognizing the limitations of evidence-based, rationalistic decision making, Weiss (1998b) thus promotes the enlightenment use of evaluations (also known as the conceptual use) in decision domains as a means to improve government programs and policies. Enlightenment is defined as "gradual sedimentation of insights, theories,

concepts and ways of looking at the world" (Weiss, 1977, p. 535). This view assumes that evaluation information affects policy decisions in more subtle and indirect ways over time, becoming a part of discursive dialogue, hence the "new common wisdom" in policy arena (Weiss, Graham, & Birkeland, 2005, p. 13). Mark, Henry and Julnes (2000), too, endorse this conception of evaluation, arguing that the purpose of evaluation is to assist deliberations in democratic institutions to improve social programs and policies for an ultimate goal of social betterment. They believe that evaluations can change assumptions, beliefs, expectations, and perceptions about programs and policies over time and ultimately influence policy decisions (Chelimsky, 2006).

Evaluation in Low and Middle Income Countries

The debate about the role and utility of evaluation for decision making remains front and center across the international development landscape. In the closing decades of the last century, evaluation systems and practice have expanded globally to contexts outside of the global North. In light of American entrepreneurship, Northern-based aid organizations have promoted evaluation as a potential decision-making tool in LMICs (Furubo, Rist, & Sandahl, 2002; Carden & Alkin, 2012; Khan, 1998). Yet many scholars argued that developing country governments do not consider evaluation beneficial for their information needs as it is considered a donor-driven activity (Conlin & Stirrat, 2008; Hay, 2010). These organizations' conceptualization of evaluation (its definition, uses and purposes) has significant implications for the value of evaluation in developing country decision-making contexts. Although much has been said about ECB in domestic, micro-level, organizational settings (e.g., Cousins, Goh, Clark, & Lee, 2004; King, 2002; Garcia-Iriarte, Suarez-Balcozar, Taylor-Ritzler, & Luna, 2011; Nielsen, Lemire, & Skov,

2011), a comprehensive understanding of the issues in a developing country governmental context is missing. This section attempts to illustrate various aid organizations' approach to evaluation and its perceived consequences for the fate of evaluation in the global South.

Background of evaluation capacity building. A search for literature focusing on evaluation in LMICs results in evaluation capacity building (ECB) reports and studies. The *evaluation capacity building (ECB)* literature provides the most traditional approach to promote evaluation as a valuable decision-making tool in LMICs. ECB activities over past several decades have been largely in response to a growing public concern for increased accountability of aid effectiveness and use of taxpayer money (Merryfield, 1985; Carden & Alkin, 2012; OECD, 2010; World Bank, 2002). In 2002, the Asian Development Bank (ADB), the World Bank, the European Bank for Reconstruction and Development (EBRD), and the African Development Bank (ADB) signed a joint statement to declare that monitoring and evaluation systems are significant in helping the donor community show effective results (World Bank, 2002). A few years after, over a hundred representatives from several donor and recipient countries, as well as bilateral and multilateral agencies, convened in Paris to discuss the present and future of development assistance. Segone (2008) thinks that the Paris Declaration of 2005 was a landmark agreement between donors and recipient countries that emphasized mutual accountability for effective development results and efficient use of funds. The Declaration demanded intensive efforts to strengthen evaluation capacity in developing countries. The flurry of ECB activities in the global South is the byproduct of this

commitment to improve recipient country ownership of and accountability for effective development results (OECD, 2006).

The responsibility to build and strengthen evaluation capacity in LMICs is primarily vested in Northern-based bilateral and multilateral aid organizations (Dabelstein, 2003). Studies drawn from the international development evaluation literature suggest that the Organization for Economic Co-operation and Development's (OECD) Development Assistance Cooperation (DAC) Network on Development Evaluation, consisting of 23 member states and development agencies, and the United Nations Children's Fund (UNICEF) are the major actors in ECB efforts. The DAC Network is an international forum for evaluation specialists and managers to combine their capacity to increase development effectiveness by high quality evaluations. The most recent survey with DAC Network members indicates that the members together produce 600 evaluations of development assistance per year and spend US$ 5.1 billion on evaluations (OECD, 2010). The same survey also indicates that half of the responding agencies have an explicit mandate in their evaluation policies for building evaluation capacity in partner countries (OECD, 2010). Among DAC members, the World Bank has the biggest budget of US$ 31 billion for evaluations, housing the largest evaluation office within the development community (World Bank, 2002, 2004). ECB has been a formal mandate of the Bank's Operations Evaluation Department (OED) since 1986, which has provided support for ECB in thirty-four countries in Africa, East Asia, Latin America, the Middle East, and South Asia (World Bank, 2004). The Bank allocated US$ 1.1 billion (i.e., 4.9% of OED's total spending) on ECB activities in 2003. In short, the Northern-

created and based aid community plays a leading role in funding ECB efforts and driving

the agenda for evaluation systems and practice in LMICs.

Definitions of evaluation capacity building. Although evaluation capacity

building (ECB) is a fairly new concept in the developing country context, the available

research is growing both in volume and importance. A substantial number of scholars

have proposed various definitions and probed the boundaries of ECB in the developing

country context, with the first major study of Boyle and Lemaire (1999). Many widely

cited and used definitions of evaluation capacity include the following:

> The human capital (skills, knowledge, experience, etc.) and financial/material
> resources, and evaluation practice to the actual 'doing' of evaluation. Evaluation
> practice refers to the definition of the evaluation, the research design, and the
> execution of the evaluation activity, that is, implementation, results, and impacts
> on specific public policy. This practice is only possible if you have the supply of
> "hardware" (in other words, evaluation capacity). (Boyle, Lemaire, & Rist, 1999,
> p. 5)

> Evaluation capacity refers to an organization's ability to bring about, align, and
> sustain its objectives, structure, processes, culture, human capital and technology
> to produce evaluative knowledge that informs ongoing practices and decision
> making in order to improve organizational effectiveness. (Mackay, 2002, p. 14)

> Evaluation capacity development refers to strengthening or building M&E
> [monitoring and evaluation] systems in borrower countries in order that a situation
> is attained where M&E is regularly conducted and utilized by the countries
> themselves- governments and civil society. In other words, M&E are tools, and
> ECD is the process of strengthening country-based systems to conduct and use
> M&E. (World Bank, 2004, p. 1)

Over the years, the need to build and strengthen evaluation capacity in LMICs has

resulted in an array of approaches and guidelines prescribed by researchers and

practitioners to improve decision making. Although findings on the integration of

evaluation into countries' decision domains are only beginning to appear in the literature,

the studies that do exist regarding ECB explicate a series of premises that underpin

researchers' thinking about establishing value of evaluation in decision-making

processes. These premises are: 1) evaluation leads to evidence-based decision making; 2)

a thorough understanding of supply and demand for evaluation will help build sustainable

evaluation systems, and 3) country people need to be involved during the ECB process

for greater buy-in for evaluation.

Evidence-based decision making. A review of ECB literature suggests that

evaluation is believed to promote evidence-based decision making that contributes to

good governance and the achievement of Millennium Development Goals (MDGs)

(Segone, 2008, 2009; Picciotto, 2003, 2011; World Bank, 2002, 2004). Davies (1999)

defines evidence-based decision making as "a policy process that helps planners make

better informed decisions by putting the best available evidence at the center of the policy

process" (as cited in Segone, 2008, p. 27). Further, Adrien and Jobin (2009) provide a

definition for good governance:

> Good governance means the manner in which power is exercised in the
> management of a country's economic and social resources for development […]
> Good governance has many desirable characteristics: it is participatory,
> consensus-oriented, accountable, transparent, responsive, effective and efficient,
> equitable and inclusive, and follows the rule of law. (p.103)

Researchers have repeatedly reported the merit of evidence-based decision making under

the banner of good governance as opposed to opinion-based policies that result from

"untested views of individuals or groups, often inspired by ideological standpoints,

prejudices, or speculative conjecture" (Segone, 2008, p. 27). Scholars believe that

monitoring and evaluation capacity will improve efficiency and effectiveness of public

policies by providing technically sound and relevant evidence to policymakers (Mackay,

2009; Picciotto, 2009). Sound evaluation evidence will help policy makers demonstrate

that they have achieved the citizens' goals in a cost-effective manner, using available

resources efficiently (Giovannini, 2009; Mackay, 2009). This helps improve the quality of decision making and strengthens accountability and transparency in governance, hence citizens' trust in government (Segone, 2008).

These researchers tend to privilege post-positivist scientific evidence with its emphasis on quantification. Available research on ECB commonly cites the significance of quantitative indicators in creating objective, credible, economical evidence for improving decisions (Rist, Boily, & Martin, 2011; Segone, Sakvarelidze, & Vadnais, 2009). Some scholars suggest that the dominance of economists and economics in development evaluation might have promoted the production of this kind of evaluation evidence (Mathur, 2009; Riddell, 1999). As a result, Northern-based organizations' ECB strategy focuses largely on building a methods-first evaluation capacity with improved quantitative databases in borrower countries (Bamberger, 1991). For example, the World Bank (2002) reported spending US$ 4 million annually to strengthen statistical capacity in LMICs. In light of studies reviewed here, it is safe to conclude that this ECB approach favors the use of quantitative evidence to ascertain some degree of objectivity in decision making (Segone & Pron, 2008).

The need for quantitative evidence for better decisions is tied to many scholars' conceptualization of evaluation. ECB researchers associated with aid organizations use the term *evaluation* interchangeably with *performance-based budgeting, results-based management,* and *financial accountability* (see Mackay, 2002, 2009; OECD, 2006; Picciotto, 2011; World Bank, 2004). Results-based management (RBM) has especially become a catch-phrase over time, conjointly and repeatedly used with the term evaluation. United Nations Development Group (2011) defines RBM as follows:

RBM is a management strategy by which all actors, contributing directly or indirectly to achieving a set of results, ensure that their processes, products, and services contributing to the achievement o the desired results (outputs, outcomes, and higher level goals or impact). (p. 2)

Based on the conceptual fusion of these terminologies, the learning and accountability functions of evaluation are commonly cited (e.g., Mackay, 2002; OECD, 2006; Rist, Boily, & Martin, 2011). Yet there is a tendency to first ensure financial *accountability* towards donors (see World Bank, 2004; UN, 2011) (e.g., step by step supervision of public expenditures) then ensure *improvement* in decision making (Schiavo-Champo, 2005). In short, researchers suggest that evaluation evidence based on financial indicators will help policy makers in LMICs demonstrate that they have achieved the citizens' goals in a cost-effective manner, using available resources efficiently (Giovannini, 2009).

Components of evaluation capacity. Researchers and evaluation practitioners alike have engaged in a continuing effort to better understand the nature of ECB at the country level. Despite the conceptual pluralism in what constitutes evaluation capacity at a national level (Rist, Boily, & Martin, 2011), researchers embrace a supply-demand model to understand the nature of evaluation capacity (see Table 3).

Table 3. A Demand and Supply Framework for Evaluation Capacity Building

		Evaluation Demand	
		Strong	*Weak*
Evaluation Supply	*Strong*	High evaluation capacity, high utilization	High evaluation capacity, limited utilization
	Weak	Limited evaluation capacity, high utilization of studies produced	Little evaluation capacity, little utilization of evaluation studies

Note. Adopted from Boyle, Lemaire, and Rist (1999).

Boyle, Lemaire, and Rist (1999) report that evaluation demand and supply impact

utilization of evaluation in decision-making processes. Following the lead of Boyle and

Lemaire (1999), many researchers proposed models to build sustainable ECB in

developing countries. Several of these models are comprehensive and notable.

First of all, OECD (2006) and UNICEF (2010) take a systems approach and

conceptualize ECB as being composed of three levels: 1) individual (i.e., knowledge and

skills to conduct and use evaluations), 2) organizational (i.e., institutional environment,

policies and legal arrangements), and 3) enabling environment (i.e., information

management systems, evaluation policy, accountability mechanisms) (see Figure 2).

Both organizations contend that a deeper understanding of contextual factors at the

country, regional, and global levels will help formulate better ECB efforts.
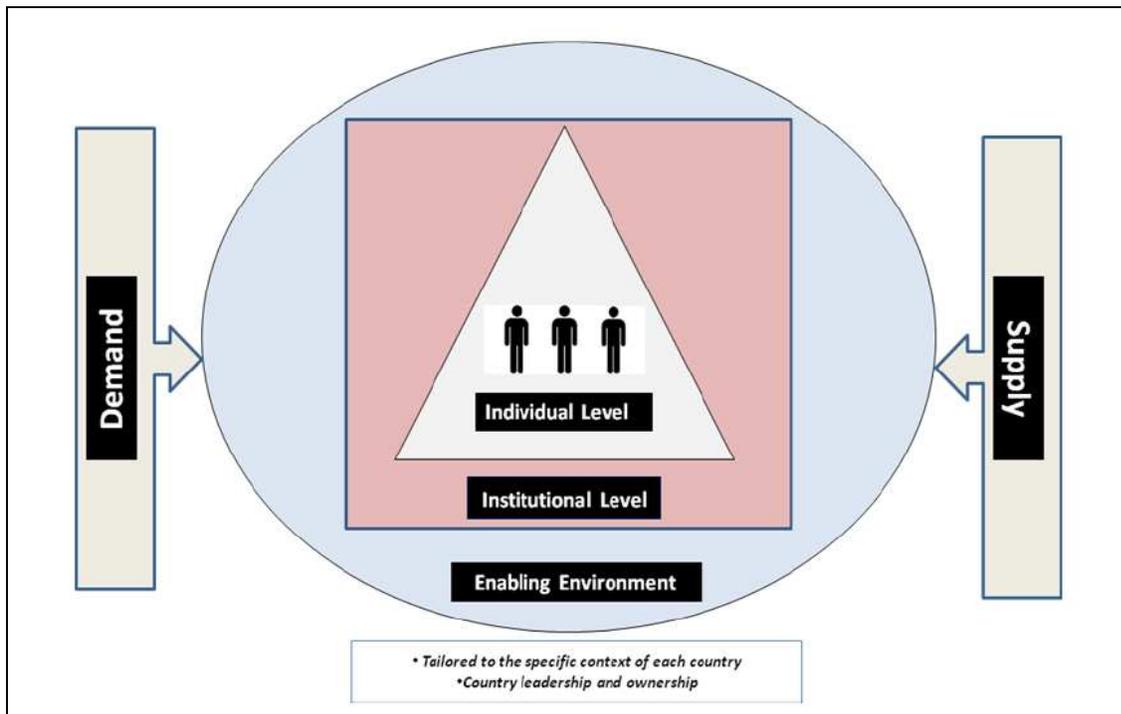
Figure 2. UNICEF's approach to building evaluation capacity (2010, p. 8).

Similarly, Rist, Boily, and Martin (2011) talk about four dimensions of evaluation capacity: institutional capital, human capital, technical capital, and financial capital. Each of these dimensions has sub-sections for supply and demand (see Figure 3). Institutional capital refers to a legal M&E framework, accountability requirements, incentives or sanctions for good performance. Human capital refers to analytical skills, training opportunities, and communities of practice. Technical capital refers to knowledge management, information systems, databases, statistical capacity, data collection and analysis, and dissemination systems. Financial capital refers to financial resources to develop appropriate M&E systems.
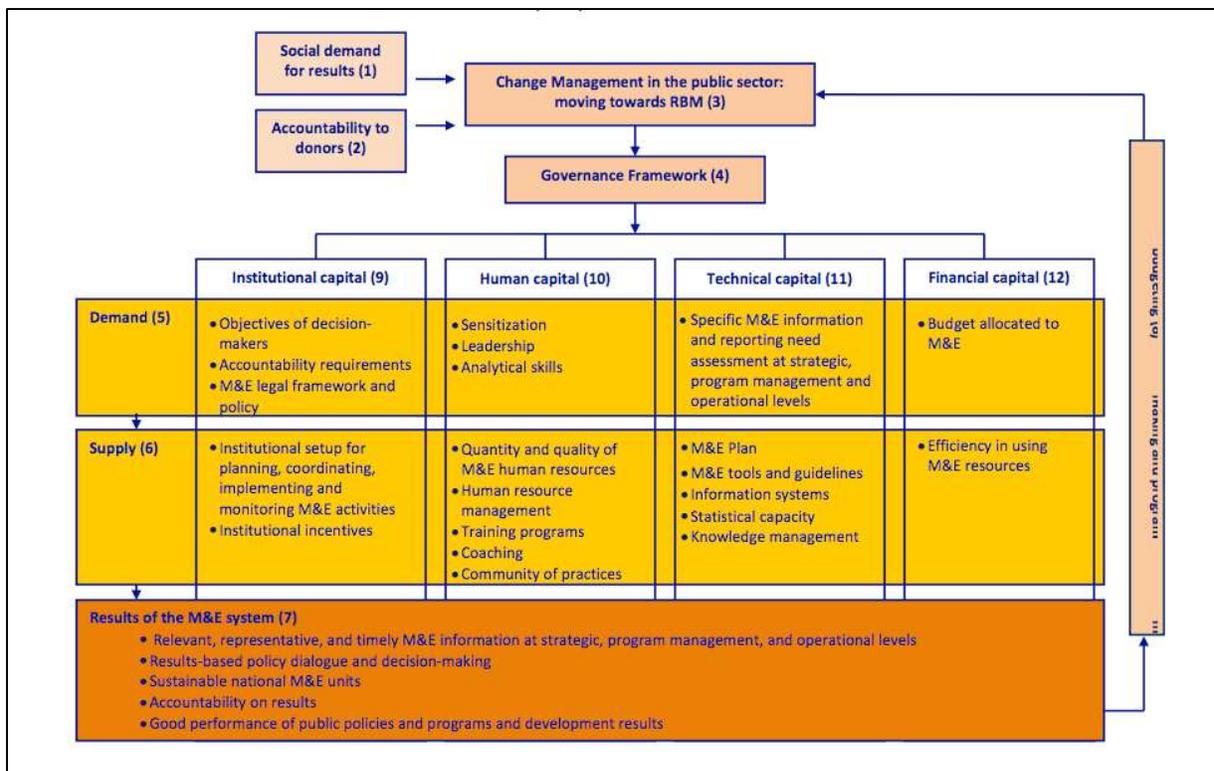
Figure 3. Evaluation capacity assessment (Rist, Boily, & Martin, 2011, p. 6).

Furubo, Rist, and Sandahl's (2002) examination of different evaluation cultures in 21 countries provides some insights into the enabling environment for evaluation

capacity. Researchers suggested that democracy, freedom of speech, freedom of press, and rationalistic decision making are prerequisites for full integration of evaluation into the decision-making process (cf. Chelimsky, 2006). The authors also contend that a strong background in the social sciences and the existence of a welfare state, which requires information about the costs and outcomes of programs, help understand the value of program evaluation and thus fully integrate it into the decision-making process.

In short, demand and supply for evaluation are thought to be the major determinants for evaluation capacity and utilization of evaluations at the LMIC level. On the demand side, researchers draw attention to existing evaluation culture at the country ministries, presence of communities of practice, incentives, mandates, and evaluation champions (i.e., individuals and organizations who are eager to promote evaluations) (Mackay, 2009; OECD, 2006; UNICEF, 2010; Horton, 2003).

Table 4. Determinants of Evaluation Culture in the Developed and Developing World

Indicator	United States	Typical Developing Country
Technology	Well-developed, sophisticated computer systems and databases that allow for methodological rigor	Limited hardware, and lack of access to large datasets
Scope	In-house, mainstream, built-in evaluations of federal programs (e.g., GAO)	Limited or almost no evaluations of government programs and policies
Finance and control	Federally or state funded	Mostly donor-funded
Who conducts evaluations	Internal (i.e., government agencies) or external (i.e., universities or consulting companies)	Mostly foreign consultants, donor agencies
Time perspective	Longitudinal and ex-post	Almost no longitudinal evaluations, which tend to end with the completion of project.

Note. Adapted from Bamberger (1991, p. 330).

On the supply side, studies focus on knowledge management systems, existing

knowledge of evaluation tools, models, and techniques, training and formal education

opportunities, funding, and evaluation standards (UNICEF, 2010; OECD, 2009). Almost

two decades ago, Bamberger (1991) compared the evaluation systems in the United

States and developing countries to describe the differences in evaluation supply (see

Table 4), which partially explains the large differences between the levels of

development of the field of practice in different regions of the world.

Despite their clearly identified models, aid organizations encounter many

challenges in building ECB in LMICs. Limited demand for evaluations, an inadequate

legal framework to support M&E systems, lack of recognition of evaluation as a

professional career, low quality information systems, lack of technical skills, and

resistance to results-based management in governments remain as challenges to building

capacity in countries (Mackay, 2009; Rist, Boily, & Martin, 2011). As a result,

researchers have explored different incentives or sanctions to increase supply of and

demand for evaluation evidence to be used for decision making (e.g., Mackay, 2009;

Schiavo-Champo, 2005). Some favor conferences, workshops, and training opportunities

to increase awareness about the utility of monitoring and evaluation evidence for decision

making (World Bank, 2004; UNICEF, 2010). Research also emphasizes the importance

of building knowledge management systems and establishing a legal framework to

mandate planning, conducting, and using evaluation information (Rist, Boily, & Martin,

2011). Mackay (2009) summarizes donors' current view on incentivizing the M&E

capacity in the developing world (see Table 5). He proposes "carrot/sticks/sermons

approach" to strengthen demand for developing and utilizing M&E systems in

developing countries (p. 180). Mackay asserts that, depending on the country's existing

evaluation culture, a combination of incentives can be provided.

Table 5. Strengthening Demand for Evaluation in Developing Countries

Carrots	Sticks	Sermons
• Provision of awards, prizes or high-level appraisal of good or best practice M&E. • Provision of budget related incentives or additional funding to ministries and agencies to conduct M&E to improve performance.	• Enact laws, decrees, or regulations mandating the planning, conduct, and reporting of M&E • Withhold funding from ministries that fail to conduct M&E.	• Organize seminars and workshops to raise awareness of M&E systems. • Organize conferences on good M&E practices, and explain what is in it for the countries.

Note. Adapted from Mackay (2009)

In addition, researchers consider regional and national evaluation organizations as

a critical part of the enabling environment for strengthening supply and demand and key

in sustaining evaluation practice by creating awareness about the utility of evaluation

systems, engaging actors in dialogue and evaluation activities (Ba Tall, 2009; Mackay,

2009; Segone & Ocampo, 2006; Piccotto, 2011). Scholars argue that the proliferation of

evaluation organizations (e.g., International Development Evaluation Association,

International Organization for Cooperation in Evaluation, African Evaluation

Association, Latin American and Caribbean Network for Monitoring, Evaluation, and

Systematization, and so on) as well as evaluation training programs for development

specialists (International Programme for Development Evaluation Training and Shanghai

International Program of Development Evaluation Training) contributed to the ECB

efforts worldwide (Lundgren & Kennedy, 2009).

Involvement of developing country governments in ECB. A review of ECB

literature suggests that Northern-based organizations tend to strategically involve partner

country representatives in ECB efforts in order to increase ownership and utilization of evaluation evidence in decision making. Researchers have recognized that it takes time, persistence, and patience to build countrywide monitoring and evaluation systems (UN, 2011; World Bank, 2004). Since the countries lack sufficient resources and demand to develop such systems (Mackay, 2009), the success of ECB efforts depends on "steady and sustained support by international donors" (Schiavo-Champo, 2005). As a result, Northern-based aid organizations propose continued involvement and financial support to achieve high-quality ECB, credible evaluation systems, and likely utilization of evaluations in decision making (Mackay, 2002). Yet a growing body of studies reports the success of meaningfully involving partner countries in the process of ECB (Estrella et al., 2000). Some researchers have recognized that capacity development is "an endogenous process of change, strongly led from within a country, with donors playing a supporting role" (OECD, 2006, p. 7). Thus, donors work with the developing country officials throughout the process of ECB to help countries understand the merit of evaluation function and articulate their policy problems and decisions (OECD, 2009; Horton, 2003).

Throughout the last decade, the intention to involve partner countries in ECB efforts has peaked with UNICEF's increasing emphasis on country-led evaluation systems (CLES) that cede control from donors to countries in evaluating the effectiveness of development goals (Segone, 2008, 2009). CLES is described as evaluations that the developing country both leads and owns by determining "what policy or programme [sic] will be evaluated; what evaluation questions will be asked; what methods will be used, and what analytical approach will be undertaken" (Segone, 2009, p.24). CLES is based

on respecting national ownership and leadership, political processes, priorities and strategies, as well as cultural values in order to effectively serve countries' information needs (UNICEF, 2010). Researchers and practitioners argue that CLES seek to give voice to diverse stakeholders within countries through participatory processes, and creates a "virtuous cycle of better public policy" (Ba Tall, 2009, p. 119; Adrien & Jobin, 2009). Notwithstanding the bourgeoning literature on the significant of CLES, its impact on strengthening the value of evaluation as a decision making tool is yet to be empirically investigated.

Strengths and weaknesses of ECB in the global South. The major strength of Northern-based ECB efforts is its promotion of evaluation as a potential decision-making tool in developing countries (Carden & Alkin, 2012; World Bank, 2004). Available research suggests that the Northern-based aid community has helped demystify evaluation practice in the developing world (Mackay, 2009; UNDP, 2011). Moreover, despite their small scales, Schiavo-Champo (2005) suggested that ECB efforts conducted by the World Bank's Operations Evaluation Department (OED) contributed to the development of public monitoring and evaluation capacities in some countries (e.g., Uganda and Egypt).

In addition, ECB researchers have provided a promising approach to understand the main components of evaluation capacity (supply and demand) (Horton, 2003). Their research can be utilized to fully address the conceptual pluralism regarding the ways to build evaluation systems in developing countries so that ineffective practices will not be adopted repeatedly. One significant illustration of mainstreaming and standardizing evaluation practice in non-Western settings is the introduction of international evaluation

standards (see Russon & Russon, 2005). Another example is the *Glossary of Key Terms in Evaluation and Results-based Management* created by OECD's DAC Network (2002) as a modest step to harmonize evaluation capacity related activities around the world. Turkey has a Turkish version of this glossary jointly prepared with the Turkish Development Cooperation (Kocaman & Guven, 2008).

Similarly, many ECB studies repeatedly report the use of a diagnostic analysis of the countries' existing evaluation systems (Mackay, 2009; OECD, 2006; Rist, Boyli, & Martin, 2011; World Bank, 2004) that help modify the donor support for ECB and its priorities. OECD (2006) recognizes that the capacity development is by no means only technical transfer of knowledge and skills "from North to South"; the most critical part of the process is "the main impulse from within" the recipient country to own and drive the process (p. 15). Available research suggests that ECB efforts target at creating realistic, country-specific action plans for ECB by understanding the current public management culture, identification of M&E practices in government departments, the influence of such activities in decision making, key players and their information needs, and the role of international donors (Mackay, 2002; Schiavo-Champo, 2005). Thus, researchers have recognized that one-size-fits-all or cookie cutter approach will not accommodate the specific country context (OECD, 2006). Finally, researchers increasingly favor the participation of country people in the ECB process to increase their chances of understanding the merit of evaluation, articulating policy problems, and solutions, and supporting their sense of ownership of evaluation processes (Dabelstein, 2003; Estrella et al., 2000; Horton, 2003).

Despite its strengths, the literature on ECB efforts in developing countries is as strong as its weakest link. Many scholars question the scale of ECB initiatives across diverse regional and country contexts as opposed to a single organizational level (King, 2010; Sridharan & De Silva, 2010). Some fear the uneven distribution of evaluation capacity within the country across institutions and organizations (Kumar, 2010; Bamberger, 1991). Thus, the widespread success of ECB still remains as a challenge (Carden, 2010).

First and foremost, there is a growing body of research investigating the ways to increase country ownership of evaluation processes and findings, but relatively little attention has been given to an equally important topic: how can Northern-based ECB activities become a part of national decision making beyond development assistance? Capacity building does not guarantee that evaluation will become a routine part of daily decision-making processes (Sanders, 2003). Researchers suggest that ECB activities are designed to mainly evaluate development programs, and most ECB efforts focus on evaluating donor programs against donor criteria for donor needs (Carden, 2010; Hay, 2010; Sridharan & De Silva, 2010; Bamberger, 1991). While Bamberger (1991) calls the donor imposition in the field of evaluation "cultural imperialism" (p. 337), Picciotto (2007) describes it as "business-as-usual" whereby resources to enhance evaluation capacity at the country level remain embedded in donor agencies; thus, donors' imposition for one-sided accountability continues (p. 512). While it is well understood that aid organizations' continued involvement in ECB is needed to develop effective evaluation systems in some LMICs due to a lack of resources, what remains unclear is an understanding of how donors' need for demonstrating "value for money" (OECD, 2010)

will translate into mainstream evaluation function for national decision making at the country level (Sridharan & De Silva, 2010; Kumar, 2010).

In addition, the existing literature converges to argue that a number of developing countries have strengthened their evaluation systems and practice, yet little is known about how LMICs conceptualize evaluation for their own national information needs (UNDP, 2011). In the most thorough study to date on evaluation cultures across twenty-one countries, Furubo, Rist and Sandahl (2002) recognized that evaluation approaches and models were disseminated from the larger aid organizations and added (cf. Patton, 2001):

> Late comers have adopted these ideas, perhaps to show that they also subscribe to the modern and rational public management school of thought. But the conclusion here is that adherence to these ideas in most cases has been mainly lip service. (p. 17)

In other words, when the demand is not from within, then the evaluation activities have remained at the level of "lip service" (Mackay, 2002, p. 95) whereby the developing countries unquestionably accept the evaluation systems without understanding the utility or practicality of proposed systems for their governance (Bamberger, 1991).

This is mainly due to ECB activities' focus on financial accountability; performance-based budgeting, and monitoring rather than on learning-oriented evaluations to improve national programs and policies (Ba Tall, 2011; World Bank, 2004). Some scholars have recognized that, due to ECB efforts, most evaluation activities in LMICs have been confined to the finance sector and budgeting initiatives (Dabelstein, 2003; Bamberger, 1991). Schiavo-Champo (2005) describes these types of structures as "evaluation ghettos" (p. 9) where in-house evaluations in selected government departments are not systematically and effectively connected to other actors in

governance, and they are not necessarily carrying out comprehensive, systematic

evaluations. Thus, evaluation culture has yet to permeate into other administrative

structures and the entirety of political culture (Kumar, 2010).

This focus on monitoring and accountability as opposed to evaluation and

learning merits further investigation and discussion, particularly in the context of

improved decision making. Researchers have claimed that some developing countries

(e.g., Uganda, Costa Rica) are over-monitoring programs to achieve quantitative

performance indicators and demonstrate accountability (see World Bank, 2004; UNDP,

2011). For example, World Bank's *Self-Evaluation Report* (2004) provided findings for

ECB activities in Uganda and Egypt that received high-intensity support from the Bank.

In Uganda, the report indicated that evaluation equates monitoring and concluded that,

"Uganda has too much monitoring and too little evaluation. The feedback loop from

findings to policy is still weak and in some sectors nonexistent. More evaluation is

needed" (p. 54). Researchers, however, argue that monitoring may not easily reveal

inaccessible and mysterious internal processes of a program to grasp what must be going

on inside the *black box* (Booth & Lucas, 2001). While discussing the various purposes of

evaluation, presented earlier in this chapter, Patton (2012) also commented on "the

marriage of monitoring and evaluation" in developing countries (p. 124). He seems to be

suspicious of the utility of installing performance monitoring systems in international

contexts, and argues that:

> More often, as soon as accountability mandates are introduced, and they're
> introduced early and authoritatively, the tail wags the dog, and everyone focuses
> on meeting accountability demands, effectively undercutting the learning and
> improvement agenda, and limiting managerial willingness and capability to take
> risks that might attract opposition or resistance. It's not enough to create results-
> oriented monitoring systems. An organizational culture and climate must be

created to support the appropriate and effective use of such systems. (p. 124) Rogers (2005) endorsed Patton's position, arguing that accountability systems do not necessarily satisfy decision makers' information needs due to their heightened focus on compliance and oversight.

Moreover, research that equates performance-based budgeting and program evaluation may not accurately capture the nuances of social programs and policies and may improperly prescribe evaluation models and systems that focus only on financial accountability with quantitative indicators of measurable services without paying attention to the quality of such services (Bamberger, 1991). As a result, ECB activities' focus on financial monitoring and accountability may reduce the potential of program evaluation's learning function to improve decision making (Dabelstein, 2003; Ba Tall, 2011). Researchers suggest that monitoring-focused financial indicators may not be appropriate to capture the nuances of social programs, and may easily lead to corruption (Bamberger, 1991).

Some researchers argue that ECB efforts' focus on building statistical capacity to produce rigorous evidence for decision making is narrow. Kawakami and colleagues (2008) questioned the political rhetoric of scientific rigor because of which evaluation adheres to strict standards and assume that the strict definitions of validity can become an element of control in the hands of an evaluator that may potentially lead to cultural misinterpretations. Borrowing the term "epistemological racism" from Scheurich and Young (1997), the authors claim that validity is a "white, majority Western thought" (p. 202) that is likely to compromise the interests of underrepresented groups because it limits the "questions asked, theories considered, designs selected, measurement strategies employed" (p. 202) (cf. Kirkhart, 2005). As such, narrow constructions of validity can

further marginalize non-majority groups' ways of knowing and fail to appreciate the cultural nuances in method selection. The authors state that:

> Less frequently considered is the fact that the core assumptions underlying inquiry—the evaluative questions asked, the voice in which they are expressed, the choice of success indicators, the operationalization of positive outcomes, etc.—may lose their integrity in translation to new cultural contexts. (p. 203)

Furthermore, even the growing interest and investment in joint evaluations and country-led evaluation systems do not seem to remedy the lack of in-country demand to use evaluations for decision making. A survey of DAC Members indicated mixed evidence regarding the utility and success of partner country involvement in evaluations. Agencies that responded to the survey indicated that the involvement of partner country stakeholders in evaluation activities has not gone beyond collecting data (OECD, 2010). The report concluded that despite an increasing interest, stakeholder involvement in designing evaluations, selecting evaluation questions, and disseminating results has been limited. Dabelstein (2003) commented on the process of joint-country evaluations that are geared towards increasing local ownership:

> Common to most of these evaluations is that the developing countries have played a minor – if any – role in their planning and execution. Rarely are the recipient countries involved until an evaluation scheduled by the donor is initiated and most often the recipient government is involved in providing information to the donor but not in the analyses and final assessment of performance. Evaluation programmes [sic] are prepared in response to agency needs around lesson learning and accountability, and are geared to the planning and programming cycle of the agency, not to the planning cycles of partner countries. (p. 367)

Last but not least, the diversity of ECB activities funded and supported by different donors may be contributing to underutilization of evaluation in decision making. Different donors build and strengthen different capacity for evaluation of their own development programs due to differences in approaches to ECB (OECD, 2010). Without harmonization of systems, developing countries may end up with fragmented systems of

capacity that do not work together. An evaluation of the implementation of the Paris

Declaration revealed that the donors have difficulty in harmonizing their priorities and

capacities with the recipient countries (OECD, 2009; Carden, 2010). The World Bank

(2004) is aware that harmonization of ECD practices across donor agencies (or the

development community) is a prerequisite to reducing the burden on developing

countries. The lack of harmonization also reduces the chances of shifting the focus from

the donors' needs to country needs (Carden, 2010).

Taken together, based on their theoretical assumptions, ECB scholars appear to

grant the control to Northern-based donor agencies in building evaluation systems and

practice for better informed decision making in LMICs. Their approach is driven by time

and resource limitations in LMICs that justify donor control in creating systematic and

valid evaluation systems. It is unclear, however, how LMICs view the role and utility of

evaluation for their national decision making if donors exclusively control it. Also, little

is known about national context that informs the conceptualization and utility of

evaluation for national decision making. Without that context, national decision makers

may be poorly informed about the value of evaluation for their particular information

needs beyond development assistance. To this end, gaining better insights into the

countries' existing decision-making and learning systems, as well as country context may

prove useful in understanding the value of evaluation in decision making.

Indigenous Evaluation Systems and Practice in the Global South

> *Until fairly recently, [evaluation] theory was positioned almost exclusively in a*
> *White, male, heterosexual, academically educated, Eurocentric majority context*
> *that was not understood as "cultural" due to environmental universality; that is,*
> *the invisibility of majority privilege. (Kirkhart, 2010, p. 402)*

Recently, calls to explore the ascribed value of evaluation from the LMIC perspective have increased in the evaluation literature. In 2010, the *Forum* in the *American Journal of Evaluation* focused on evaluation field building in South Asia to elevate the discussion about the state of program evaluation in developing countries and discuss the opportunities to develop evaluation as a profession outside of the global North. The forum problematized the dominance of Northern-based and -created organizations in evaluation in developing countries and suggested that as long as these organizations are preeminent in driving the evaluation agenda, evaluation will remain weak in developing countries (Carden, 2007, 2010). Carden (2007) calls this "the real evaluation gap" where evaluation remains weak in LMICs due to Northern-created institutions' dominance in the field (p. 220). Thus, Carden (2010) and Hay (2010) advocated defining the value of evaluation by and for the country stakeholders themselves for their own information needs and developing their indigenous systems and practice from the bottom up. By indigenous, researchers mean:

> Evaluators in different countries around the world are developing their own [evaluation] infrastructures to support their endeavors as well as their own preferred theoretical and methodological approaches. (Chelimsky & Shadish, 1997, p. xii)

This relatively new approach to understanding the value of evaluation from the developing country perspective is beyond building capacity, embracing a broader context for highlighting the utility of evaluation for domestic decision making.

An indigenous, Southern approach to evaluation seeks to eliminate the influence and control of Northern-based organizations in the field of evaluation in order for program evaluation to penetrate developing countries' national decision-making processes beyond development assistance. A growing body of researchers has demanded

an answer to the question of whose interests are being served by particular evaluation

agendas, models, and questions (Sridharan & De Silva, 2010; Smith, 2012; LaFrance &

Nichols, 2008; Hopson et al., 2012). Some other researchers contend that Northern-based

ECB has been merely a technical transfer, hence inadequate to remedy the growing

weakness in and limited public sector evaluations in the developing world (Carden, 2007;

Hay, 2010). Indeed, Hay (2010) argues that the enthusiasm among the donor community

in building evaluation capacity "camouflages…the declining or stagnant state of

evaluation in South Asia" (p. 223).

Studies drawn from development evaluation literature report that state and non-

state actors in some developing countries are still widely suspicious and cynical about the

value of evaluation in decision-making processes because they do not see the utility of

donor-driven evaluation activities for their own information needs (Hay, 2011). In

addition, scholars submit that developing country nationals might view evaluation as a

Western, imperialist notion that subjugates and marginalizes local knowledge and styles

of decision making (Bhola, 2003; LaFrance, 2004; Smith, 2012; Rai, 2001; Merryfield,

1985). Menon (2011), the head of Evaluation Office in UNDP, once argued:

> The extensive discussion on evaluation capacity development among international
> development practitioners appears to assume that the approaches and systems of
> evaluation in the bureaucracies of international development partners should be
> replicated in national systems. We are not convinced that this is fully correct.
> Many of us in multilateral and bilateral development agencies have invested in
> fine-tuning evaluation systems that address our specific organizational and
> governance needs. However, these systems do not mesh naturally with national
> systems. Nor are they always effective in addressing the accountability concerns
> of citizens of developing countries regarding development cooperation within
> national development. (p. 13)

As a result, scholars argue for an evolution of indigenous evaluation cultures and

capabilities by and for developing country nationals to self-determine their

understandings of social problems, responses, and ways to create useful knowledge that

will contribute to the country's own national decision making and solution of social

problems (Carden, 2007; Smith, 2012; Kawakami et al., 2008). These scholars do not

suggest changing the definition of evaluation, but rather contextualize the practice in a

way that is more attuned to local realities. They advocate for responsiveness to the

historical, political, and cultural context of the country and welcome the

reconceptualization of evaluation, the need and capacity for evaluation, and its valued

purposes within that context (Hay, 2011). They eventually advocate for eliminating the

monopoly of Western institutions and values over the field in developing countries and

testing all evaluation theories and methods from scratch based on their own information

needs so that evaluations can meaningfully contribute to decision making (Hay, 2010;

Carden, 2010).

  The study by LaFrance and Nichols (2008) provides the most thorough example

of grounding the value of evaluation in a particular context and for a particular audience.

Employing a national sample of Native Americans, the researchers developed a bottom-

up *Indigenous Evaluation Framework (*IEF) based on reinterpretation of program

evaluation from the Native American perspective in order to encourage evaluators to

critically rethink context within the parameters of the culture of the evaluand. Kirkhart

(2010) described IEF as "a culturally specific theory [built] *from scratch*, through

inductive information gathering from members of the intended context" (p. 406, italics

added). To them, culture is the central element that permeates and defines the terms of all

dimensions of context. Culture defines and mediates what constitutes a social problem

and genuine knowledge; how merit and worth are understood; what responses and

alternatives are appropriate, and how the communities' wellbeing can be advanced (LaFrance, Nichols, & Kirkhart, 2012). IEF invites evaluators to reexamine their epistemological, methodological, and axiological assumptions, which define the legitimacy of evaluation conducted in Native settings. They note, "Evaluators [should] step in rhythm with the community rather than setting their own pace" (LaFrance, Nichols, & Kirkhart, 2012, p.69) in their pursuit of trustworthy information, which is itself culturally determined. This indicates that cultural context defines the value of evaluation.
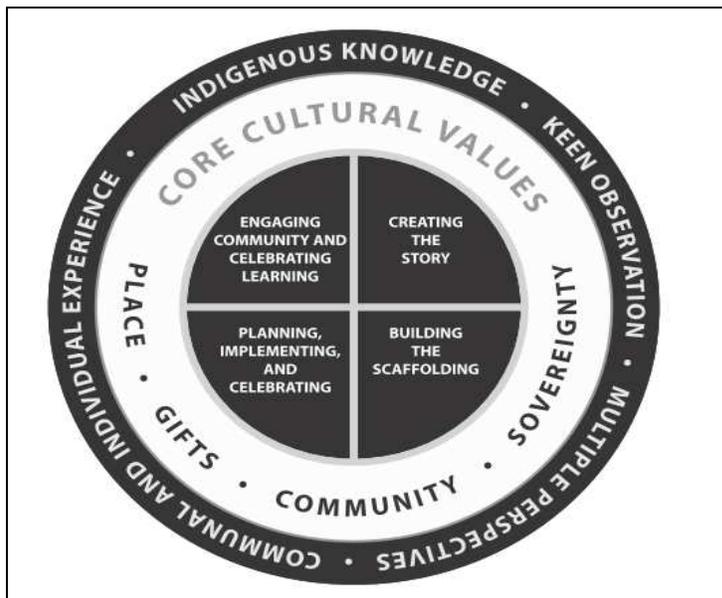
Figure 4. Indigenous Evaluation Framework (LaFrance & Nichols, 2012, p. 63).

Although LaFrance et al. (2012) give a warning about the limitations of IEF for evaluation practice outside of Native American contexts, the issues and concerns IEF raises illustrate and provide good ideas about extending the debate on the value of program evaluation in the developing country context (see Figure 4 for the depiction of LaFrance and Nichols' model). IEF highlights the significance of understanding context

as a force in shaping the development and nature of evaluation practice in different countries (Rist, 1990; Fitzpatrick, Sanders, & Worthen, 2004; Neirotti, 2012; Hay, 2010). Dramatic differences in context for evaluation across countries highlights the problem of adopting Western notions, methods, models, and theories of evaluation in different country settings because evaluations will be embedded in multiple layers of historical, political, social, and cultural realities (Fitzpatrick, 2012; Nevo, 1982). In addition, IEF demonstrates how the historical record of evaluation in developing countries can dramatically affect the value of evaluation in the eyes of developing country people, which merits further investigation.

Another notable example of the changing role and value of evaluation in a national context is from Denmark. Dahler-Larson and Schwandt (2012) examine how the political cultural context informs and continues to shape the nature of the role and utility of evaluation for decision making in Denmark. Researchers submit that, on the one hand, the historic underpinnings of the Danish welfare state (i.e., social justice, collectivism, equality, redistribution, transparency, dialogue) has created a participatory and democratic evaluation practice to obtain citizen input, accommodate multiple viewpoints, balance consensus, and compromise to improve programs. On the other, the European Union's (EU) push for competitive policymaking has recently introduced a new direction for evaluation in Denmark that focuses on quantitative studies, especially in the fields of education, health, and employment so that Danish government can sustain its competitiveness within and outside of the EU context. The authors argue that evaluation and historical, institutional, and cultural context are co-constructed.

In addition to examples from the global North, there are a few investigations of indigenous, home-grown evaluation cultures in the developing world. A prominent example of context-specific value of evaluation for context-specific needs is from Latin America where the development of evaluation as a discipline goes back to the 1990s. In 2012, an entire issue of *New Directions for Evaluation* examined the impact of sociopolitical context on the development of evaluation as a field of knowledge in the continent (Kushner & Rotondo, 2012). Scholars discussed how evaluation is perceived as an opportunity to deepen democracy, increase citizen participation, and provide equitable social results for all in the region where many countries are in post-conflict stages (Neirotti, 2012). Latin American researchers argued that the role of government; citizens' view of the government; and the conditions of civil society have affected the nature of public policies and evaluation function thereof (Neirotti, 2012; Martinic, 2012). Scholars also discussed how Paulo Freire's concepts (participation, empowerment, citizen agency, indigenous knowledge, and marginalization) seem to have an impact on the discourse around evaluation in Latin America (Guendel, 2012). The forum suggested that evaluation practice in Latin America tends to give priority to community values, local knowledge and priorities; thus, participation has become the "Latin American stamp" on evaluation approaches and models (Kushner & Rotondo, 2012, p. 1). In sum, many Latin American countries (e.g., Brazil, Mexico, Costa Rica) are thought to develop their home-grown evaluation cultures and systems based on their own contextual needs to improve their decision making (Mackay, 2002; UNDP, 2011). The National System for Results Evaluation of the Public Administration in Colombia, the National System of Evaluation

in Costa Rica, and program evaluation built into the Brazilian Pluriannual Action Plans at federal, state, and municipal levels are all illustrations of this.

Although the literature on indigenous understanding of evaluation in the developing country context is in its infancy, this approach still provides clear and justifiable points. Above all, these scholars place evaluation practice and final decision making in the hands of developing country people. They advocate for giving the ownership of the evaluation systems and practice to the national governments, eliminating the preeminence of the international institutions in the field so as to enhance national decision making. An indigenous approach to grounding the value of evaluation in the country context welcomes non-fixity of evaluation practice, and demands unlearning and relearning the role and utility of program evaluations from the country people's perspective. Thus, the biggest strength of their studies resides in their ability to disrupt and challenge the status quo in the field of evaluation.

Taken together, this body of literature adopts a critical lens to explain and justify the reasons and consequences of using evaluation in-country decision-making processes in the global South. An indigenous evaluation approach gives prominence to historical, social, cultural, and political context to understand the value of program evaluation indecision-making processes. Rejecting Northern-based metanarratives and essential meanings in the practice of evaluation, these authors imply that there is not a unified and fixed meaning of evaluation practice. The scholars advocate for indigenous, contextually congruent evaluation systems and practice in developing countries to improve national decision-making processes.

Conclusion

In this chapter, I reviewed relevant bodies of literature with regard to their premises and implications for exploring the value of evaluation in a developing country context at large. I argued that the value of evaluation as a decision-making tool will require an understanding of contextual factors that shape and continue to inform the utility and relevance of evaluation for in-country decision making.

Efforts to build evaluation capacity in low and middle-income countries are undertaken in the presumption that evaluation is vital to governmental decision-making (Boyle, Lemaire, & Rist, 1999). Established scholars in the field have long argued that evaluations should be integrated into decision-making processes and practices at the domestic and international level because it adds social and economic value to the work of organizations or institutions (Russ-Eft & Preskill, 2009). Boyle, Lemaire, and Rist (1999), in a landmark study of ECB, discussed the waves of countries that sought to build evaluation capacity and institutionalize evaluation function in public policymaking. The first and second wave countries almost exclusively from the global North including the United States, Canada, Sweden, Denmark, Great Britain, the Netherlands and Norway attempted to integrate evaluation into their governmental decision making for the purposes of program improvement and public accountability starting 1960s. The countries from the global South including Korea, Indonesia, Colombia, and Zimbabwe, were exemplified as the third wave that blended features of accountability and improvement under the leadership of the World Bank. Based on these examples, the authors present evaluation as an effective tool of governance, almost unquestionably, and argue that evaluation is treated as a "positive national investment" (p. 3) in all of these

country cases hence the trend to build governmental evaluation regimes around the world. Despite the results-based management framework's promising potential to integrate evaluation skills and knowledge into governmental practices around the world, the resulting success in many countries has not been satisfying (Bourgeois & Cousins, 2013).

Evidence from a wide range of evaluation and policy scholars and practitioners converges to suggest that indigenous evaluation systems and practice in the developing country context may provide effective guidance on exploring the utilization of program evaluations embedded in Turkey's decision context and responsive to its political culture. Demand and supply models promoted by the Northern-Based ECB approach also provide valuable insights into the factors that affect the role and utility of program evaluation in educational practices in Turkey. In general, I believe this review contributes to the emerging notion of indigenous evaluation systems and practice in contexts outside of the global North. A picture is emerging that building indigenous evaluation systems is vitally important to learn to improve national decision making at the country level. Chapter 4 will exemplify the context of evaluation in the country case of Turkey, particularly in relation to education decision domain.

Chapter III: Methodology

This chapter describes the methodology for exploring the value of evaluation as a decision-making tool in Turkish educational decision domain from the perspectives of key stakeholders – government officials, academics, and civil society representatives – and key documents authored by these stakeholders. The perspectives gained from this study will not only contribute to our understanding of the role of evaluative activities in governmental decision-making processes in a developing country, but will also inform the process of evaluation capacity and field building efforts in the global South at large. Additionally, the study findings make theoretical contributions to the field of evaluation by highlighting the contextual dynamics at play in the global South that might affect the international expansion of the field. Although I do not claim any external generalizability, there is no obvious reason to argue that the findings could not shed light on other country cases (Maxwell, 2005).

The research is guided by the following sub-research questions:

1. To what extent and in what ways is evaluation utilized in the Turkish education decision domain?

2. How is evaluation conceptualized in the Turkish education decision domain?

3. What is the need and capacity for evaluation in the Turkish education decision domain?

4. What contextual factors facilitate or hinder the value of evaluation in Turkey's educational decision-making?

Guided by the critical and constructivist paradigms, the design of this study gives primacy to the views, opinions, and lived experiences of the study participants in

exploring the value of evaluation in Turkish educational decision context. The first question is geared towards mapping the current evaluation agenda – institutionalization of evaluation function, roles and responsibilities of M&E units, purposes, uses, and methodology of evaluation – in educational decision-making process, exploring how this agenda reflects on and projects upon the actual, expected, and imagined trajectory of evaluation practice in Turkish governmental sphere. The second question explores the many meanings and margins of evaluation as a concept and a form of inquiry in Turkish educational domain. The third question is designed to not only reveal demand for evaluation, but also assess existing and potential capacity to undertake evaluation for education public programs and policies. The last question aims to identify the contextual factors associated with the role and utilization of evaluation as a decision-making tool. Critical orientation of this study is evident in decolonizing the practice of evaluation by privileging the voices of stakeholders from a low and middle-income country whose perspectives are missing in the mainstream literature. As such, the study makes a geographical contribution to empirical literature on evaluation utilization, cross-cultural expansion of evaluation, evaluation capacity building and field development.

The remainder of this chapter further develops the methodological approach by first describing the origin of the study, and then elaborating on the research design and methods, data collection process and analysis procedures. The chapter concludes with ethical considerations including research reflectivity and the principle of intersubjectivity. Data collection and analysis protocols, and timelines are also illustrated.

Origin of the Study and Pilot Research

This section briefly outlines the personal journey that motivated me to study the value of evaluation in my native country Turkey. Maxwell (2005) argues that, "It is important to recognize and take into account of the personal goals that drive and influence your research. Attempting to exclude your personal goals and concerns from the design of your research is neither possible nor necessary" (p. 19). Maxwell (2005) nicely captures the power of personal interests in carrying out research studies, which is highly relevant to my case. I follow the literature on evaluation capacity and field building efforts in the global South and have yet to find any mention of evaluation activities in Turkey, which I consider crucial to the global expansion of the field. As outlined in the first chapter, formal public program and policy evaluations are rare in Turkey, in an era when evaluation is promoted as an essential tool for decision-making and a vital element in new public management reform.

My interest in exploring evaluation's value as a decision-making tool in Turkish stems not only from my national background and aspirations but also from personal and professional experience with evaluation and policy studies in other country contexts. So far in my career, I have designed and implemented evaluations for non-profits, school districts, state government agencies, and various collaborations and foundations across the United States, Europe, and Africa with programs ranging from small pilots to international initiatives. These experiences soon made me realize that the stakeholders' understanding, perceptions, attitudes, and experiences related to program and policy evaluation differ dramatically. More often, the intention was to aid the decision making of any sort; thus, the utility of evaluation was judged based on its contribution to

information generation. Yet, context defined the terms of *utility* and *value*. Considering the domestic and international affairs within which Turkish decision-makers maneuver, I developed a strong interest in first investigating evaluation's value for decision making – which developed members of the global North has been preaching – in Turkish education domain without imposing contextual meanings and practice of evaluation on Turkish stakeholders.

In addition to informal conversations I have had over the years with various Turkish thought leaders, government officials, and academics, I carried out a pilot research during the summer of 2012. I conducted preliminary, informational interviews with two professors of educational administration and planning at two esteemed research universities in Ankara, as well as a formal interview with a government official at the Ministry of National Education, and a Judge at the Supreme Court of Cessations to understand the need and timing of such research. Considering the recent developments in the political, economic, social, and legal arenas in Turkey, these informants stated that my research would shed light on decision practices at the governmental level and raise awareness regarding effective decision-making procedures in Turkey.

Research Design

This research employs an exploratory, single-case study design. Since the current study places a special interest in the individual case itself – the Turkish education decision domain – without a purpose to generalize to other cases, a case study approach is suitable to shed light on a "specific, unique, bounded system" (Stake, 2005, p. 445), and unearth localized and obscure meanings as well as experiential knowledge within that situated context.

Case study approach. A case study approach is suitable for this research to provide rich and holistic exploration about contemporary issues and events over which the researcher has little or no control (Merriam, 1998, Hancock & Algozzine, 2011). Yin (1989) defines case study as an "an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used" (p. 23). According to Yin (1994), a case study design is appropriate when a what, how, and/or why question is being asked about a current-day phenomena without manipulating its context. Stake (2005) notes that case studies garner insights into "the nature of the case, its activities, historical background, physical setting, other economic, political and legal context, other cases, those informants through whom the case can be known" (p. 447). In doing so, case studies allow for converging data from multiple sources (i.e., documents, interviews, surveys, observations, etc.) and making theoretical propositions about the phenomenon without generalizing to other populations or universes.

An important aspect of a case study design is to define what the case is. Stake (2000, p. 23) describes a case in the following way:

> It can be whatever "bounded system" is of interest. An institution, a program, a responsibility, a collection or a population can be the case. This is not to trivialize the notion of "case" but to note the generality of the case study method in preparation for noting its distinctiveness.

Following the lead of Yin (1994) and Stake (2005), the current single-case study explores the value of evaluation as a decision-making tool – a contemporary phenomenon – within the Turkish educational decision context by using a variety of evidence (semi-structured interviews, documents, and field notes). The case in this study is the Turkish educational decision domain that is chronically underrepresented in the field of

evaluation, thus presenting a unique context worth exploring and analyzing. This case

presents a critical opportunity for such an exploration due to the recent developments in

educational context that permit evaluation to play a visible role as discussed in the first

chapter (see Figure 5 for the case illustration). In the present study, educational domain is

defined as the bounded physical and conceptual space where public policies, programs

and projects pertaining specifically to education are imagined, deliberated, developed,

negotiated, enacted, and enforced. As such, the domain comprises of a collection of

ministries, academic institutions, and civil society organizations that have a stake in

educational decisions, hence an influence on deliberating, determining, and shaping such

decisions.
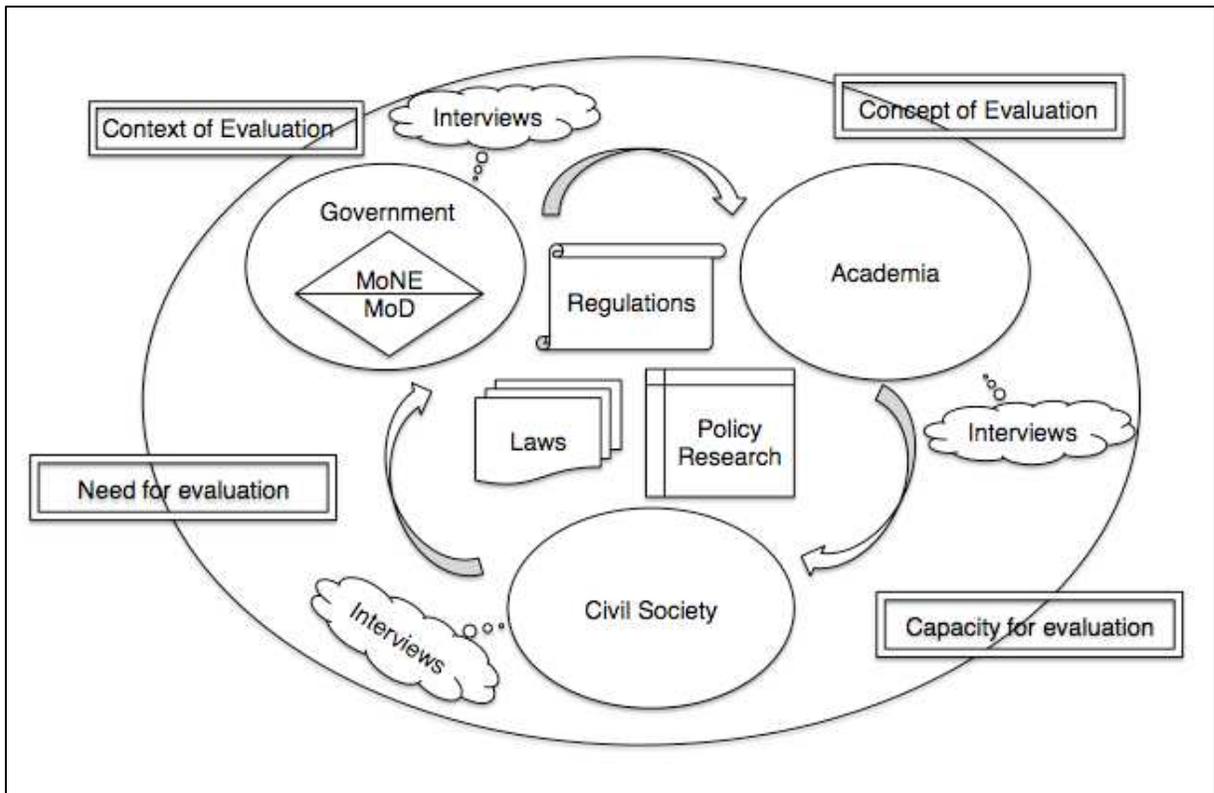
Figure 5. The Turkish Education Decision Domain

Within this bounded system, the current case study will help develop and refine theoretical propositions about the role and utility of evaluation in educational context in a developing country. The aims of this study are heuristic, so it will focus more on exploring relationships than testing hypotheses. By doing so, this study will help refocus future investigations regarding evaluation utilization in low and middle-income countries.

Selection of sites and participants. This case study involves more than one unit of analysis whereby the attention will be also directed to subunits within the case of educational decision domain. Thus, it is an embedded design (Yin, 1994), whereby the larger and central unit of analysis is the educational decision domain consisting of government, academia, and civil society with corresponding institutions and individuals.

I used *purposeful sampling* to select the institutions and individuals within these institutions that I could learn the most from (Patton, 2002; Creswell, 2007; Merriam, 1998). As noted by Patton (2002), "the logic and power of purposeful sampling lies in selecting *information-rich cases* for study in depth. Information rich-cases are those from which one can learn a great deal about issues of central importance to the purpose of inquiry, thus the term purposeful sampling" (p. 230, *italics in original*). Because of lack of historical data and prior research, the present study specifically used *critical case sampling* to investigate the value of evaluation in educational decision domain within the institutions that are expected to have the greatest impact on educational decisions, hence the role and utility of evaluation. Hence, under the decision domain as the larger unit of analysis – which the study ultimately makes conclusions about – the selection of subunits was driven by two related criteria: (1) the degree of influence on educational decisions, including program and policy agenda setting, design, implementation, and enforcement,

(2) the degree of openness and willingness to work with a Turkish researcher who has completed her graduate degree in the global North with an extensive theoretical and experiential knowledge in the field of evaluation. As a result, the research took place at three different spheres of influence, namely governmental departments – the Ministry of National Education and the Ministry of Development to be specific – research universities and civil society organizations – think tank institutions – in Ankara, Istanbul, and Eskisehir that are actively engaged in educational research and evaluation in addition to their ongoing professional contact with educational authorities.

First, the Ministry of National Education (MoNE) and the Ministry of Development (MoD) are the two central governmental organizations that impact educational policies and programs from design to implementation. To illustrate the critical nature of these selected institutions, a brief description of the organizational structure of MoNE and MoD is needed. First of all, MoNE consists of four parts: central organization, provincial organization, overseas organization, and related institutions (see Table 6 for the organizational chart of the central administration). The main decision-making units under the central organization includes the Ministerial Office, the Board of Education, the Board of Inspection, the Board of Strategy Development, and four Deputy Undersecretaries which house main and auxiliary service units. Among these units, the Board of Education is a decision-making and scientific consultation body of the Ministry, and directly reports to the Minister. The Board develops curriculum, lesson plans, and objectives, related course materials, and seeks the Minister's approval for implementation and dissemination decisions. The Board also houses a Monitoring and Evaluation Unit

that assesses the implementation and effectiveness of curricular activities across the system. As a result, the Board constituted a critical leg of the research sites.

Table 6. The Structure of MoNE's Central Organization
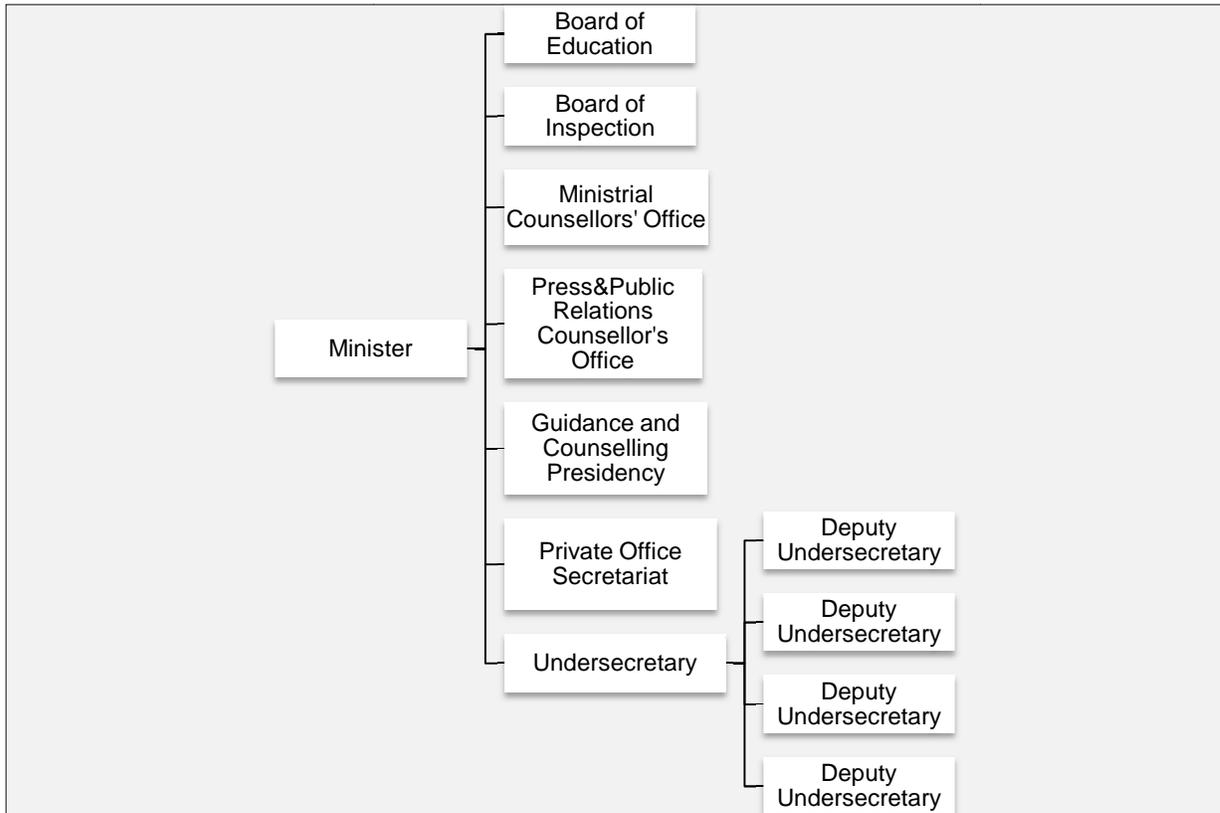
Under four Deputy Undersecretaries (see Table 7 for main service divisions; red-lined divisions have Monitoring and Evaluation Units), service units of General Directorate of Primary Education, General Directorate of Secondary Education, General Directorate of Religious Education, General Directorate of Strategy Development, General Directorate of Vocational and Technical Training, and General Directorate of European Union and Foreign Relations were selected as additional research sites due to their support in implementing educational programs and policies, designing strategies for educational targets, creating data management systems, and following the imperatives of

EU regulations – matters of importance for the purpose of current study. Thus, these units are considered critical in exploring the role and utility of evaluation.

Table 7. General Directorates at the Central Organization of MoNE

Deputy Undersecretary	Deputy Undersecretary	Deputy Undersecretary	Deputy Undersecretary
General Directorate of Primary Education	General Directorate of Teacher Training	General Directorate of Educational Technologies	General Directorate of Vocational and Technical Training
General Directorate of Secondary Education	General Directorate of Human Resources	General Directorate of Strategy Development	General Directorate of Lifelong Learning
General Directorate of Religious Education	Support Services	General Directorate of Information Technologies	General Directorate of Private Education Institutions
General Directorate of Special Education and Counseling	Legal Consultancy	General Directorate of Property and Real Estate	General Directorate of EU and Foreign Relations

Second, the Ministry of Development (MoD) is the main government body that guides Turkey's development process. The Ministry advises government in developing economic, social and cultural programs and policies, including education, and ensures the implementation of development policies and strategies though effective and efficient use of resources. The Ministry comprises the Office of Executive Assistance, Ministry Advisors, International Auditing Department, Legal Advisory Unit, Secretariat of Committees/Councils and three Deputy Undersecretaries, which provide technical support and service to the Undersecretary in development-related matters such as regional development, investment portfolios, economic modeling (see

www.mod.gov.tr/Lists/OrganizationChart/Attachments/2/KBOrganizationChartEng.pdf).

Among these service units, Education and Culture Department constituted the central

research cite at MoD that houses extensive knowledge and experience related to

evaluation activities in educational contexts. In sum, these selected governmental

departments house extensive library and archival holdings that document research

activities and policy studies of national education programs and policies.

Second, colleges of education located at research universities in Turkey constitute

second research site. Turkey houses 173 universities as of 2012 out of which 107 are

public and 66 private universities. Almost all of these universities have a School or

College of Education, which houses various departments with specialty areas. For the

purpose of this study, the Departments of Educational Policy, Planning, Administration,

Measurement, Assessment, and/or Educational Psychology as well as Curriculum and

Instruction were targeted due to their close engagement in educational matters and

evaluation activities. Specifically, large research universities in Ankara, Istanbul, and

Eskisehir house the most active educational departments in terms of the volume and

significance of the research activities and involvement with educational affairs. Once I

created a list of potential universities, I reviewed each education department's website to

select sites that are the most active in educational research and evaluation in Turkey.

Based on the volume and significance of educational policy research and evaluation

produced as well as the volume of direct consultation services provided to the Ministry of

National Education, Gazi University (Ankara), Hacettepe University (Ankara), Middle

East Technical University (Ankara), Ankara University (Ankara), Istanbul University

(Istanbul), and Osmangazi University (Eskisehir) were selected as the most critical to

provide rich insights into the educational decision making context and the role of evaluation in this setting.

Finally, civil society constituted the last but not least critical research setting. One pillar that is both a cause and an effect of Turkey's surge for development is active involvement of civil society organizations in governmental affairs either through official project and program design and implementation or informal mechanisms to inform decision-making process. For the purpose of current study, civil society organizations with a special focus on and dedicated funds for educational policy research and evaluation were targeted. Two criteria that specifically guided the selection are (1) the degree to which a civil society organization directly works with educational officials evident in consulting or project services provided, and (2) the volume of research and evaluation on current educational affairs produced evident in the existence of thematic policy briefs, periodic reports, op-ed articles, and press releases. Based on these criteria, two most active civil society think-tank organizations, located in Ankara and Istanbul, were selected as research sites.

The second unit of analysis of the current study is the individuals within these selected institutions. To determine the potential list of interview subjects, I used the institutional websites for government, academia and civil society. The following criteria were used to locate information rich cases: (1) the formal position occupied – executive directors, professors, and presidents; (2) the years of work experience in current profession, as well as knowledge of, and involvement with evaluation, and research activities and decision-making process because of the formal position they occupy – determined by professional experience noted in official CVs. Keeping in mind that the

purpose of typical case sampling is "illustrative not definitive" (Patton, 2002, p. 236), I

created a list of potential interview subjects, and I consulted with the key contacts at each

research site to recruit the subjects who were willing to spare their time for an interview.

Table 8. Study Sampling

Level	Institutions	Units	Individuals
Government	MoNE	Board of Education; General Directorates of Primary Education, Secondary Education, Religious Education, Vocational and Technical Training, Strategy Development, European Union and Foreign Relations	General Directors and/or senior officials who directly work with educational programs and policies, and have a working knowledge of evaluation and research
	MoD	Education and Culture Department	General Directors and/or senior officials who directly work with educational programs and policies, and have a working knowledge of evaluation and research
Academia	College or School of Education	Departments of Educational Policy and Administration, Measurement, Assessment, Educational Psychology	Tenure-track professors
Civil Society	Educational think-tank organizations	Education research offices	Directors and research associates

I used the point of redundancy and saturation as the criteria to decide the size of

the sample. Following the lead of Lincoln and Guba (1985), I stopped recruiting

information rich cases from the list when additional interviews started generating no new

information at which point I assumed information collected was maximized.

Data Collection Procedures

In order to harness many perspectives to bear on the original research question,

the current study employed three methods, namely semi-structured interviews, document reviews, and field notes, to converge information from multiple sources over a period of 6 months from August 2013 until January 2014 (see Table 9 for the data collection procedures). The information collected was mostly descriptive, deriving primarily from the experiences and perceptions of the key study participants – government officials, academics, and civil society representatives. The documents were used, to a large extent, to support the narrative of the study.

Table 9. Data Collection Procedures by the Unit of Analysis

Unit of Analysis	Data Collection Source	Type of Data	Method
Decision Domain	System	Structural properties (Government Plans, National Development Plans, European Union Acquis)	Documents
Government, Academia, Civil Society	Organization & Institution	Education policies and research, laws and regulations, statutory decrees, strategic plans	Documents, and Interviews
Individuals	Government Officials, Academics, and Civil Society Representatives	Perceptions of and experiences with why and how the education decision domain works the way it does, and how evaluation penetrates	Interviews, and field notes

In-depth interviews. In order to garner comprehensive and historical information about the value of evaluation, which cannot be observed directly, I conducted face-to-face individual interviews with a total of 35 information rich cases. Right after the Eid al-Fitr (i.e., Feast of Breaking) observed during August 7 - 12, 2013, I contacted the General Directorate of Educational Technology located at the Ministry of National Education's Besevler Campus, Ankara – the organization that issues permission to external researchers to contact Ministerial staff for interviews and access official documents at the

Ministry. Once I obtained the official permission letter by August 26, 2013 (see Appendix C), I started scheduling interviews first with government officials at MoNE based on the list created to explore the context within which educational decisions are made. Almost every government official contacted – except for two senior officials who did not show any interest in participating despite three attempts via first email then phone calls – responded positively to my request for an interview. While face-to-face interviews with academics and civil society representatives were scheduled primarily via email, I endeavored to meet with the potential government interviewees first face to face to describe the purpose of the research, answer any questions they had regarding the study, and to create rapport so that their concerns would be eliminated, and a trustworthy conversation would be possible. I managed to conduct 18 in-depth interviews with government officials at which point I reached empirical saturation with interviews yielding significant consistency with previously collected data.

I first obtained approval from the University of Minnesota's Institutional Review Board in June of 2013 to begin my study (IRB HSC #1306P36461). Starting mid-September 2013, I started contacting academics and civil society representatives via email to schedule interviews. While 3 academics declined my request due to time conflict, 5 never responded even after two email reminders. The rest of the potential academic participants listed and all think-tank representatives contacted expressed their interest in the study. Due to geographical separation of nongovernmental participants as opposed to one central location of government officials, scheduling interviews presented significant challenges. Despite the limited funds available to conduct the research, and recurring time conflicts with a few academics, I managed to interview 13 professors and

4 civil society representatives in Ankara, Istanbul, and Eskisehir.

The main purpose of the interviews was to investigate how government officials, academics, and civil society representatives view the value of evaluation for educational decision making in Turkey (see Appendix A and B for the interview protocols). I asked respondents to comment on current decision-making practices and mechanisms, and if, why, and how evaluation is integrated into these practices or remains on the periphery. I invited respondents to explain the opportunities and constraints existing in the decision domain that make it more or less likely for evaluation to influence educational policies and programs. Interviews helped key stakeholders co-construct a rich description of their understanding of evaluation practice and its role in decision making within their own cultural worldview and discursive practice and within the institutions they work at. In acknowledging that the study participants have had different expertise, and opinions about the given issue, the interview process provided diversity in viewpoints and experiences. Respondents represented a wide range of age and educational background, with close to equal representation of men and women (see Table 10 for characteristics of individuals interviewed).

All interviews were conducted in Turkish. A majority of the interviews (21 out of 35) were recorded with the consent of the participants. These informants were assured that their information will be kept confidential, and their names will not be attached to any comments used in the report of the findings. All of these digitally recorded interviews were transcribed in Turkish by a professional transcriptionist from Turkey to produce verbatim transcripts. I listened to every recording while I read the transcriptions to make corrections where necessary and jot down additional notes about respondents'

pauses and hedges. I managed to take notes as verbatim as possible during unrecorded interviews, and kindly ask the participants to repeat statements that appeared imminent to the study's questions to make sure they are transcribed verbatim real time. All quotations used in this study are participants' actual statements. Wherever possible, I corrected grammar and syntax of quotations for clarity without diminishing participants' representation of reality or meanings they attached to constructs of interest. Two government officials requested my interview notes for their own records. These interview transcripts were emailed after they were edited and formatted for clarity and presentation.

Table 10. Individual Interview Participants

Institutions	#Individuals Interviewed	Gender Breakdown	Highest Educational Degree Obtained (location)
Ministry of National Education	16	Female: 4 Male: 12	Bachelor: 3 (domestic) Master: 5 (domestic) Ph.D.: 8 (7 domestic, 1 overseas)
Ministry of Development	2	Female: 1 Male: 1	Master: 1 (domestic) Ph.D.: 1 (domestic, ongoing)
College of Education	13	Female: 2 Male: 11	Ph.D.: 13 (8 domestic, 5 overseas)
Educational think-tank organizations	4	Female: 0 Male: 4	Master: 1 (overseas) Ph.D.: 3 (1 domestic, 2 overseas)
Total	**35**	**Female: 7 Male: 28**	**Bachelor: 3 Master: 7 Ph.D.: 25**

Document review. A review of official policy documents corroborated evidence gathered from the individual interviews. Without accepting them as literal, documents and archives helped garner insights into governing code of conduct and behaviors in the Turkish educational decision domain and the meanings main actors attach to decision making and evaluation. These data sources also provided further insights into the climate

surrounding the decision context and helped find contradictions for further investigation (Yin, 1994). Additionally, these artifacts helped obtain data in an unobtrusive way and understand the language used by study participants when they are making decisions and engaged in evaluation-like activities (Creswell & Clark, 2007).

For the purpose of this study, paper or computerized artifacts that include public documents (e.g., policy briefs, progress reports, formal evaluation studies, minutes of meetings), internal documents (i.e., agendas, proposals, memorandums, letters), any service or organizational records (i.e., e-school, MEBBIS, educational surveys) that comprise evaluation systems were targeted to understand the existing evaluation systems and structure, as well as the literal and nascent capacity for conducting and using evaluations. In addition to documents produced and utilized by national actors, I also collected supporting, complementary documents produced by international donors primarily to understand the larger context within which Turkish education domain, specifically educational agenda setting and policymaking is being influenced.

The criteria for choosing documents and records for review were (1) the document was created no earlier than 1990 when the educational domain in Turkey was significantly restructured; (2) the document was co-created and/or co-authored by at least two key stakeholder groups – government, academia, and civil society, and donors – to ensure credibility of findings, and (3) the authors were preferably interviewed during this study to enable cross-referencing between document review and interviews. Honoring and ensuring the confidentiality and privacy of information, each respondent was requested to share any official document that they thought pertained to the current research or access service and organizational records. Based on these criteria, almost 50

official policy and legal documents were selected and reviewed for the present study (see

Table 11 for the list of reviewed documents).

Table 11. Documents Selected and Reviewed for the Study

Level	Author	Document Name	Timeline	# of Documents
Government	JDP Government	• 59th, 60th and 61st Government Plans	2002-2015	3
		• Urgent Action Plan	2003	1
		• Vision 2023	2011	1
Line Ministries	MoD	• 8th, 9th and 10th National Development Plans	2001-2019	3
	Multi-stakeholder	• 17th and 18th National Education Council Decisions	2006 and 2010	2
	MoNE	• Annual Activity Reports	2011-2013	3
		• Strategic Plan 2010-2014	2009	1
		• Program for Strategic Plan Preparation 2015-2019	2014	1
		• EARGED Reports	1997-2010	4
Laws and Regulations	JDP Government	• Public Financial Management and Control Law No. 5018	2003	1
		• The Decree No. 652	2011	1
Civil Society	SETA	• National Education System in Turkey: Structural Problems and Recommendations	2013	1
	ERI	• Education Monitoring Report	2010-2013	6
International Donors	OECD	• Education at a Glance	2008-2013	6
		• Basic Education in Turkey: Background Report	2005	1
	EU	• Turkey Progress Reports	2005-2013	9
		• National Programme for the Adoption of the Acquis (NPAA)	2003 and 2008	2
	World Bank	• Education Sector Report	2005	1
		• Secondary Education Project: Implementation, Completion and Results Report	2012	1
		• Country Partnership Strategy 2012-2015	2012	1

Reflexive field notes. One convention this study adopted is the idea that qualitative data tends to be fluid, meaning that the researcher as the main instrument can convert any information – including individual hunches and experiences – into useful data (Patton, 2002; Yin, 1994). To be specific, I, as the researcher, am a part of the data I collect; I had my own preconceptions and expectations when I entered the field that might have colored the data I collected. Hence, reflexivity has foregrounded the data collection and analysis throughout this study. To do so, throughout the research phases, I utilized a field log to describe the research setting and participants in detail, as well as any side conversations, personal observations and insights, keeping a diary of my own thinking, assumptions, experiences, and perceptions, as well as emerging issues that assisted with data analysis later. Any event and interaction that was useful to answer the research questions was considered data and systematically captured via field notes and reflexive memos.

Field notes were immediately recorded after each interview whether it was digitally recorded or not. Each field note included information about the individual characteristics of the interviewee (the institution, gender, age, educational background, and work experience) where it was possible to obtain that information either through the interview or via the individual's online CV. The field notes also consisted of personal reflections on the interview atmosphere, including but not limited to the interviewee's attitudes towards study or specific questions, body language, the interview location. The logs also included first insights on the data collected, deciphering emerging themes and constructs to assist with the data analysis later. These reflexive notes assisted in tracing the evolution of my own analytic thinking in and outside of the field by demonstrating

how the data is shaping to argue for a specific conclusion or contradict a conclusion.

At the end, all data was organized and stored in a database to increase accessibility, availability, and reliability of evidentiary information (Yin, 1994). Table 12 below summarizes the major limitations that I encountered during the data collection process and how I attempted to eliminate these challenges to obtain reliable data.

Table 12. Data Collection Methods, Limitations, and Solutions

Methods Selected	Limitations of the Method	How This Study Attended to Them
Semi-structured interviews	Researcher Bias	• I acknowledge that it is hard to mitigate the effects of researcher bias during interviews (Maxwell, 2005). However, the benefit of obtaining rich data from this process outweighed this potential limitation.
	Need to establish continuing rapport (Maxwell, 2005)	• I took the time to meet the potential informants at least once before the actual interview to explain the purpose of the study, answer any questions or concerns they have, and thank for their time and contributions in advance.
Document Review	Internal documents may not be available for an outside researcher's review	• I asked each government official contacted if he/she can share documents that are not publicly available. I promised confidentiality and privacy of information conveyed in these documents. Three officials shared some internal research documents. For the rest, I solely relied on public documents.
	Materials may be incomplete or authenticity may be questionable (Creswell & Clark, 2007)	• I used documents as a secondary data source in this study to generate new insights that were concurrently explored with the key informant interviews.

Data Analysis Procedures

Scholars contend that data collection and analysis should go hand in hand (Maxwell, 2005; Merriam, 1998). Following Maxwell's (2005) advice to "begin data analysis immediately after finishing the first interview or observation, and continue to analyze the data as long as he or she is working on the research," (p. 95), analysis of the data I collected was an ongoing process. As soon as each interview was over, I read the unfinished transcriptions to develop preliminary ideas and categories. I annotated on the margins of the transcript what strikes me as interesting and explained why I find it interesting (see Figure 6). As interview transcriptions started piling up, I started writing this time, which I call, grand memos to narrate the tentative and emerging critical connections between different interview transcripts and documents I had reviewed. The grand memos included emerging themes and concepts, my impressions and feelings about the data records and points, contradictions, similarities, surprises, and any remaining questions.
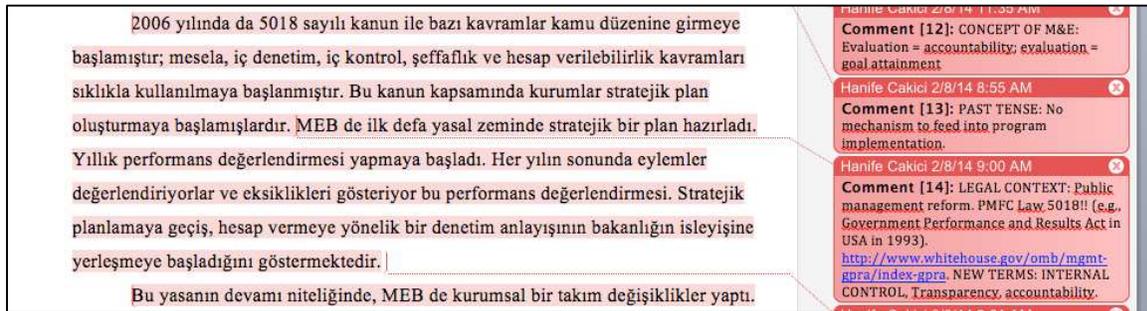
Figure 6. An Example of Annotated Memo

Once the data collection was completely over, I employed three specific methods to analyze the data: grounded theory, content analysis, and memo writing. I uploaded all interview transcriptions, all scanned documents, and field notes into the NVivo software program, which provides a structure for identifying themes and making constant

comparisons across sets of qualitative data. I first read all interview transcripts, documents, and field notes to have a general sense of the information provided over a period of six months. I once again looked for the general ideas, in-depth themes, and the tone of the conversations (Creswell & Clark, 2007). I created hyperlinks among these documents to trace the evolution of evidence. I then started the detailed analysis of data by using grounded theory approach (Strauss & Corbin, 1994) to generate categories of information through inductive, open coding (see Figure 7).
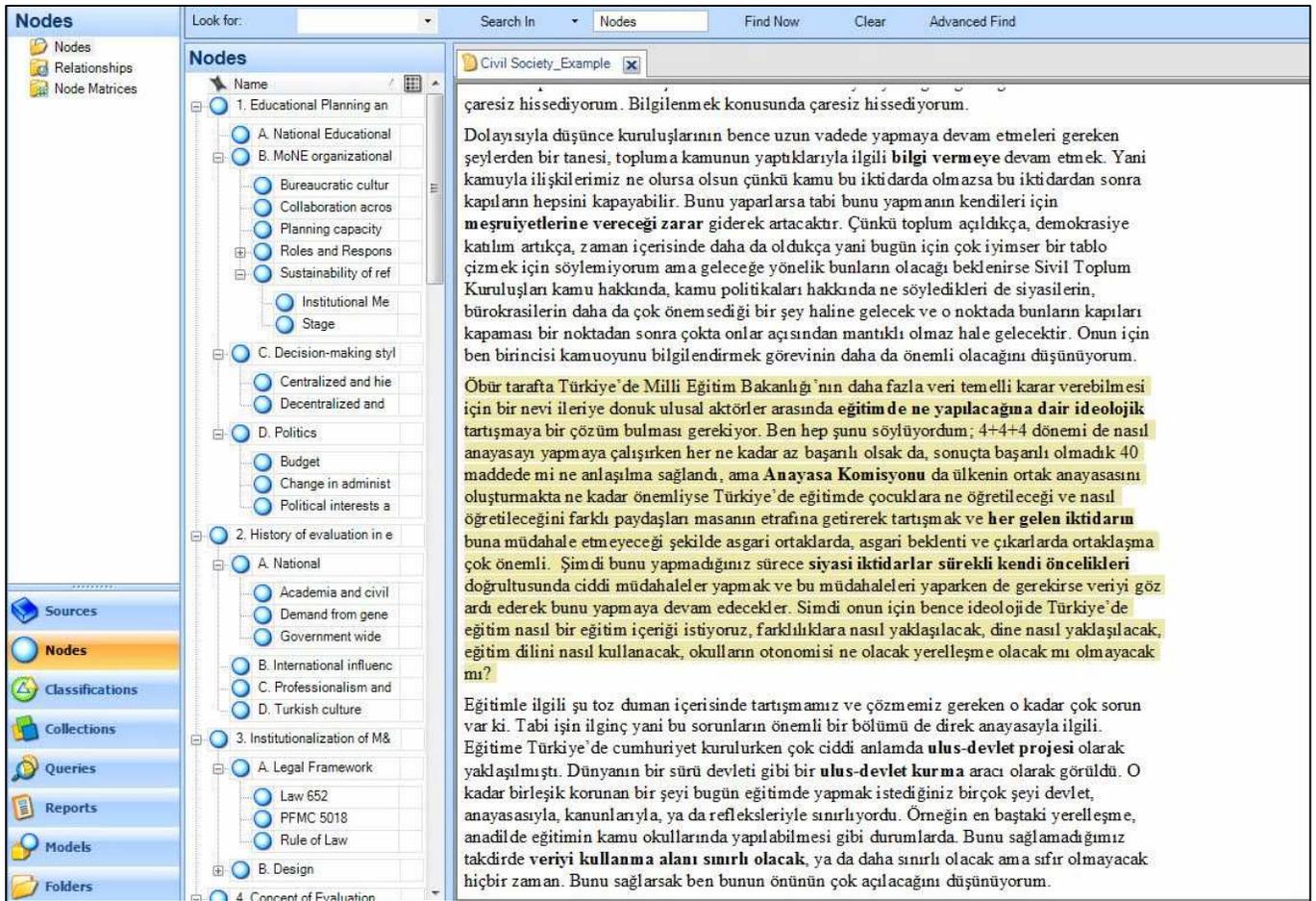
Figure 7. An Example of a Coded Interview

Maxwell (2005) makes a distinction between organizational, substantive, and theoretical categories. While organizational categories aim at sorting data into descriptive topical

groups (that could serve as the chapter headings), substantive categories seek to describe

the meaning in the data segments based on participants' perceptions. For example,

"bureaucratic culture" is an example of an organizational category acting as a conceptual

box to hold all related data, "clumsy" is an example of substantive category that captures

the meaning that some participants attached to the bureaucratic culture. I used NVivo

software to create organizational and substantive categories (see Figure 8).

Name	Sources	References	Created On
1. Educational Planning and Programmin	0	0	2/13/2014 10:43 AM
A. National Educational Philosophy	14	37	2/15/2014 11:17 AM
B. MoNE organizational structure	0	0	2/13/2014 10:44 AM
Bureaucratic culture	21	73	2/13/2014 11:46 AM
Collaboration across units	19	33	2/13/2014 11:45 AM
Planning capacity	15	42	2/20/2014 6:34 PM
Roles and Responsibilities	3	7	2/13/2014 10:44 AM
Sustainability of reforms	2	2	2/13/2014 11:46 AM
Institutional Memory	18	25	2/13/2014 11:46 AM
Stage	19	53	2/13/2014 11:46 AM
C. Decision-making style	0	0	2/13/2014 10:45 AM
Centralized and hierarchical	21	60	2/13/2014 10:46 AM
Decentralized and participatory	23	61	2/13/2014 10:45 AM
D. Politics	0	0	2/13/2014 10:46 AM
Budget	9	14	2/13/2014 10:46 AM
Change in administration	14	22	2/13/2014 10:46 AM
Political interests and power	21	75	2/13/2014 10:46 AM
2. History of evaluation in educational pla	0	0	2/13/2014 10:47 AM
A. National	0	0	2/13/2014 10:47 AM
Academia and civil society studie	12	40	2/13/2014 10:48 AM
Demand from general public and	11	33	2/20/2014 10:37 AM
Government wide	28	75	2/13/2014 10:47 AM
B. International influence and borrowi	22	76	2/13/2014 10:48 AM
C. Professionalism and expertise in t	12	23	2/13/2014 10:48 AM
D. Turkish culture	16	66	2/13/2014 10:48 AM
3. Institutionalization of M&E	0	0	2/13/2014 10:49 AM
A. Legal Framework	0	0	2/13/2014 10:49 AM
Law 652	22	52	2/13/2014 10:49 AM
PFMC 5018	13	21	2/13/2014 10:49 AM
Rule of Law	16	31	2/13/2014 10:49 AM

Figure 8. Partial Display of the Coding Scheme

I checked the position of these codes within the conceptual framework outlined before, and look for discrepant cases and arguments. Generating codes was an iterative process in that codes, concepts, and ideas have changed over time. I kept a log frame throughout the analysis of data with dates and time to describe and justify any changes in the coding tree (see Appendix D for the final coding tree). I wrote memos to create analytic connections among substantive categories accounting for all relevant data.

Analysis of interview data was aided by the content analysis of documents reviewed concurrently while coding the interviews. I attempted to narrate the relationship between interview codes and document content to partially answer my research questions. Quotations, some visuals, and potentially tables accompanied the narrative.

Last but not least, I wrote memos throughout the analysis of qualitative data. Maxwell (2005) notes that, "[Y]ou should regularly keep memos while you are doing data analysis; memos not only capture your analytical thinking about your data, but also facilitate such thinking, stimulating analytic insights" (p. 96). The memos helped capture the contextual elements that the structured data collection tools could not capture. The memos also exposed the evolution of my thinking and interpretive framework.

Triangulation, more specifically "structural corroboration" (Eisner, 1998), among these data sources was devised to achieve construct validity, and periodic conversations with government officials who are familiar with the educational context were sought to discuss emerging conclusions from the study. Corroboratory mode, converging evidence from multiple sources, has helped bring in many perspectives to bear on the original research question. I conducted progressive focusing (Stake, 2010) to check for rival explanations and negative evidence throughout the analysis.

Validity Issues

Among other things, there is one major threat to credibility of my findings: researcher bias (Maxwell, 2005). My Turkish upbringing and Western education put me in a unique position as a researcher. My values pertaining to the welfare of my fellow Turkish citizens and my acculturated expectations related to the use of evaluation may introduce an undeniable bias into the study conclusions. My lack of involvement with the study context for the last few years, however, limited my ability to understand the nuances and subtleties of the current decision-making practices. Using the checklist provided by Maxwell (2005), I endeavored to utilize two specific strategies to minimize if not eliminate the researcher bias:

Triangulation. This refers to obtaining varied and detailed information from the study participants that can negate the impact of researcher bias (Maxwell, 2005). In addition to verbatim transcriptions, I compiled extensive and detailed field notes as discussed earlier that provided rich and detail grounding for my findings. I corroborated evidentiary data from all sources of information, including informal data-gathering sources (secondary data, interview themes, personal observations, and informal discussions with the participants) (Maxwell, 2005).

Member check. This refers to obtaining systematic feedback from the study participants about the conclusions this study draws (Maxwell, 2005). Member checks were conducted with one academic and one government official to rule out the possibility of misinterpreting informants' individual responses, and identify discrepancies between my reconstructions and respondents' discourse. The member checks helped alter or affirm the story line constructed for the study.

Conclusion

   Using Maxwell's template (2005, pp. 100-101), the Table 7 below presents the planning matrix for the research design that explicates the methods, how I obtained the data to answer the research questions and summarizes potential threats to validity.

Table 13. Data Planning Matrix for the Study

What do I need to know?	Why do I need to know this?	What kind of data will answer this question?	Where can I find the data?	Potential threats for Validity
1. To what extent and in what ways is evaluation utilized in Turkish educational decision domain?	To assess the impact of existing evaluation utilization on the value of evaluation for decision making. Institutions' current experience with evaluation (and research) will have an impact on the utility of such activities. The answer to this question can help answer Question 3 as well.	Documents (paper and computerized) that represent computer programs, online portals that comprise the literal evaluation systems, and other related policy and program documents about educational decisions and practices; semi-structured interviews.	General Directorates at MoNE and MoD	**Researcher Bias**: My understanding and definition of evaluation may have an impact on whether I consider an activity or a document "evaluative" or not. I might be selecting data that mean evaluation "to me."
2. How is evaluation conceptualized in Turkish educational decision domain?	To understand what evaluation means in this case. The way evaluation is viewed (its purpose, use, and process) has implications for its utility, role, and ultimately value for decision making. For example, if stakeholders associate the term with accountability, they may not find it desirable.	Semi-structured interviews, and field notes.	Senior officers at the Ministry of National Education, Ministry of Development who are charged with making policy and program decisions; academics and researchers across the Colleges of Education and think-tank organizations in Turkey with a specific focus on research, and evaluation.	**Reactivity**: My background and education might create a social desirability bias in a way that the respondents hide their actual experiences and perspectives related to effective decision making.

What do I need to know?	Why do I need to know this?	What kind of data will answer this question?	Where can I find the data?	Potential threats for Validity
3. What is the perceived need and capacity for evaluation in Turkish education decision domain?	To assess the demand for evaluation, and internal capacity in educational context to undertake useful evaluations for decision making. The utility of evaluation is closely linked to stakeholder buy-in as well as human and material resources.	Semi-structured interviews, field notes, documents	Senior officers at the Ministry of National Education, Ministry of Development who are charged with making policy and program decisions; academics and researchers across the Colleges of Education and think-tank organizations in Turkey with a specific focus on research, and evaluation.	**Self-report bias**: The study participants' existing knowledge of or experience with evaluation activities may be limited or quite varied. This may hinder their understanding of the "need" for evaluation. They can under or overestimate the need and capacity if they do not know about evaluation.
4. What are the contextual factors that hinder or contribute to utilization of evaluation for decision making in Turkey?	To assess the impact of the external environment (socio-economic conditions, foreign relations, cultural norms, governmental goals, etc.) on the value of evaluation for decision making. Turkey's global aspirations and relationship with the Northern-based organizations and countries will have an impact on the value of evaluation.	Semi-structured interviews, and field notes.	Senior officers at the Ministry of National Education, Ministry of Development who are charged with making policy and program decisions; academics and researchers across the Colleges of Education and think-tank organizations in Turkey with a specific focus on research, and evaluation.	**Lack of long-term involvement**: My one-site, cross-sectional study does not allow for going beyond the respondents' accounts at the time of the interview to uncover the relationship between the context (broadly defined) and the ascribed value of evaluation.

Chapter IV: Turkey's Educational Background

Today, the most important and the most productive duty of all of us is the national education affairs. We have to be absolutely successful in this area and we shall. This is the real salvation of a nation.

Mustafa Kemal Atatürk - The Founder of the Republic

Turkey serves as a perfect example of the postcolonial phenomenon of hybridity: East meets West, Europe meets Asia, secularism coexists with Islam while modernist conceptions in the form of Turkish nationalism exist side by side with postmodern ones, involving different ethnic identities seeking affirmation, greater political power or autonomy, even by armed struggle, if necessary, as has been with the Kurds.

İnal & Akkaymak (2012, p. xii)

Located between Southeastern Europe and Southwestern Asia, bordering eight distinct countries, most of which are facing serious political and economic turbulence at the time of this writing, Turkey is often cited as a success story in her unique region with increasing economic prosperity, young demographics, and a renewed role in bilateral development assistance (see Figure 9). Thanks to the country's transcontinental location and resulting geopolitical advantage, full membership to large supranational organizations (i.e., the United Nations, NATO, OECD, G-20, the Council of Europe, and Organization of the Islamic Conference, to name a few) should not come as a surprise, except for the formal membership request to the EU in 1987 and official candidate status since 2005. Housing 75 million people with a median age of 29 whose 72% live in urban locations with a gross domestic product of US$735 billion and US$15,000 per capita income, Turkey currently stands as an upper middle-income country and the 16th largest economy in the world (World Bank, 2013a), yet its position on the scale of development is often contested (Gök, 2011). With one of the highest populations below age 15 among

OECD countries (OECD, 2013), the Turkish education system with its processes and outcomes occupies a critical place in both domestic and international arenas.

Figure 9. Turkey's Geographic Location (CIA Factbook, 2013).

This section lays the groundwork for the educational decision making context in Turkey. The quantity and quality of the education system, as often explored by national and international researchers in terms of educational attainment and achievement rates, pedagogical practices, curriculum and instructional design and implementation, and teacher qualifications at classroom and school levels, are beyond the topic of this study. This section instead focuses on actors, processes, and underlining motivators of educational governance, planning, and programming at a macro level. To do so, the section benefits largely from the official documents reviewed for this study in making an introduction to the agenda status of evaluation in the education decision domain.

Educational Governance

Turkish national education system is grounded on the basis of a series of laws and regulations regulating the principles of education. Today, the Constitution of the Republic of Turkey (1982), laws on education and training, government action plans, national development plans, recommendations of the National Education Councils, and

the EU Acquis provide the foundation for MoNE's policies and programs and underscore the underlying pillars of contemporary Turkish national education (see Figure 10); that is, Turkish education shall be national, democratic, secular, scientific, and modern in order to improve the welfare of Turkish citizens, enhance national unity, and contribute to the social, economic and cultural development of the country (Ministry of National Education, 2002). Within this legal framework, the Constitution (1982), the Basic Law of National Education No. 1739 (1973), and the Law on the Organization and Duties of MoNE No. 3797 (1992) regulate the general framework and principles of education for different levels and functions, and determine the framework for policies and strategies implemented through national and government plans and programs.
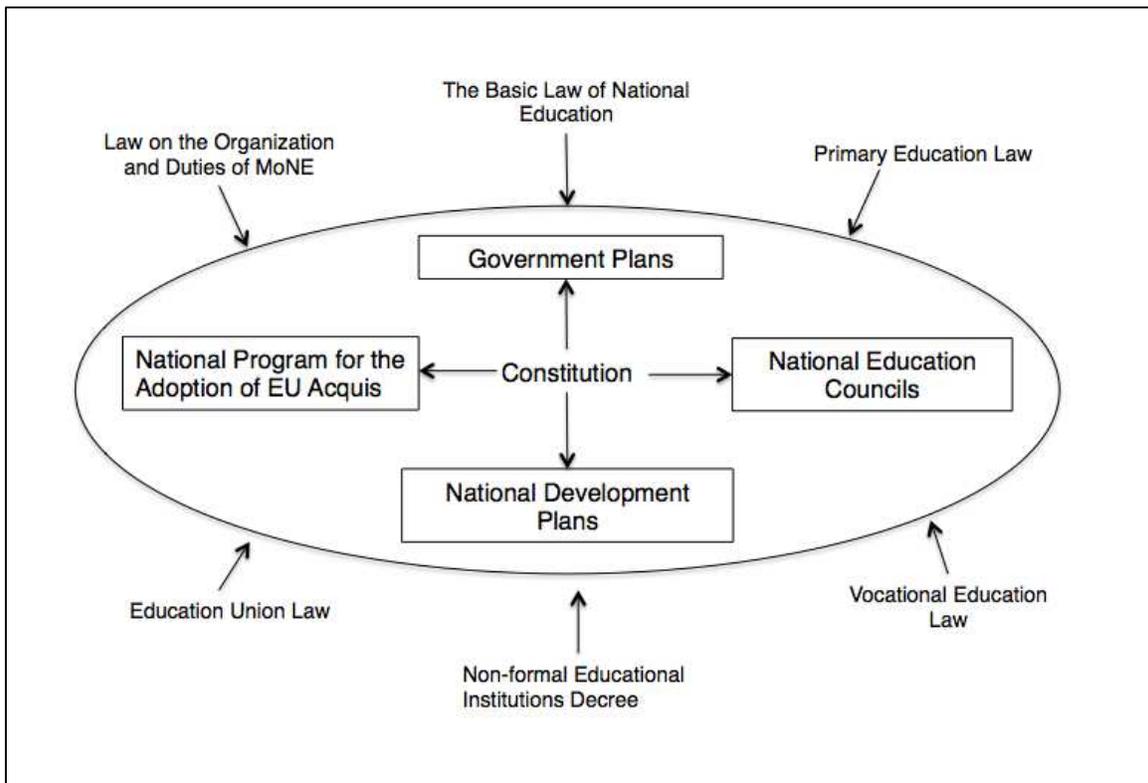
Figure 10. Legal Framework Regulating Turkey's Education System

Turkey houses one of the most centralized education systems among OECD countries (World Bank, 2005; Eğinli, 2010) (see Figure 11). As a part of the movement of westernization and modernization, Mustafa Kemal Ataturk, the founder of the Republic, launched a series of education policies (also known as Kemalist revolution) during the first 10 years of the Republic that provided the foundation for today's educational system in Turkey. The first major Kemalist reform was to bring all education institutions under the control of the state in 1924, endowing MoNE with the responsibility for all educational endeavors to ensure that every Turkish citizen receives a worthwhile education in accordance with his/her talents, interests, and aptitudes, while gaining necessary knowledge, skills, behavior, and habits (Ministry of National Education, 2002).
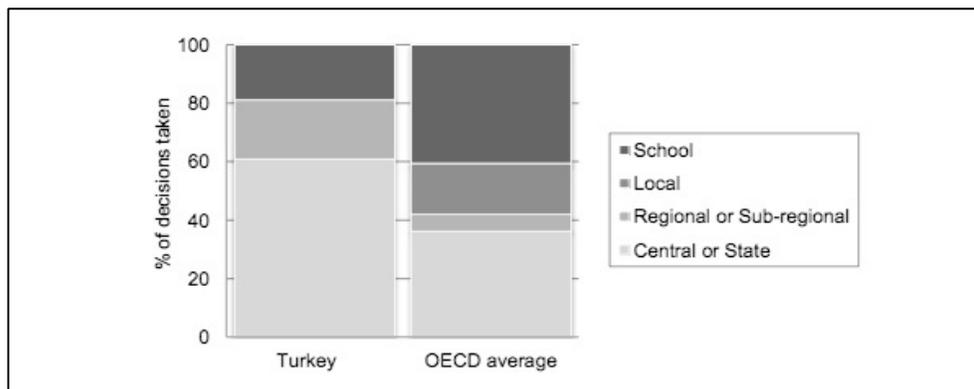
Figure 11. Percentage of decisions taken on public lower secondary schools at each level of government (OECD, 2013, p. 16).

Specifically, the Law of Unification of Instruction No.430 enacted in 1924 centered the power, control, and authority of managing, designing, implementing, and monitoring educational services across 81 Turkish provinces at MoNE. According to the Basic Law of National Education, MoNE determines all matters related to curriculum, teacher appointments, student examinations, education finance, school buildings, and classroom

materials including textbooks, and this tendency to centralize decision-making remains intact to this date (Gershberg, 2005; Keser-Aschenberger, 2012).

Allocation of financial resources exemplifies Turkey's highly centralized education system. The central government controls education funds and expenditures for pre-primary, primary, and secondary education through public funds allocated and appropriated by the national budget. The public education spending increased from $21 billion in 2006 to $31 billion in 2011, and the share of MoNE's budget in GDP increased from 2.13% to a 3.02% during the AK Party era (see Table 14). Yet, the ratio of 3% for K-12 education is still below the OECD-recommended average of 3.9% (OECD, 2013).

Table 14. Public Spending on Education in Turkey

Year	Ministry of National Education Budget to GDP Ratio	Higher Education Council and University Budgets to GDP Ratio	Total Education Budget to GDP Ratio
2002	2.13	0.71	2.84
2003	2.24	0.75	2.99
2004	2.21	0.79	3.00
2005	2.29	0.81	3.10
2006	2.18	0.78	2.96
2007	2.53	0.78	3.31
2008	2.41	0.77	3.18
2009	2.88	0.97	3.85
2010	2.57	0.84	3.41
2011	2.63	0.93	3.56
2012	2.73	0.91	3.64
2013	3.02	0.97	3.99

Note. Adopted from Çelik & Gür (2013, p. 164)

Schools enjoy limited autonomy in financial management by receiving in-kind contributions from parents and private organizations. International organizations (EU, World Bank and UNICEF to be the prime actors), nongovernmental organizations and private institutions also contribute to available educational funds in Turkey. Centralized

funding for education is considered as a barrier to eradicating the financial disparities across schools and regions, especially for the ones with lower socio-economic status (OECD, 2013).

In addition to top-down decision-making structure inscribed into the system since its inception, the Turkish education decision domain is often characterized as highly politicized and bureaucratic, precluding the Ministry's capacity to design and implement effective and responsive policies and programs (Gür, 2011). Gür and Çelik (2009) once observed: "Underlining many problems in national education is the tendency to turn educational policymaking into a partisan arena whereby ideological attitudes are displayed" (p. 12). Echoing this concern, the urgent need to restructure the Turkish education system has been one of the most frequently mentioned recommendations in national development plans (Şişman, 2011). Restructuring and reformulating the national education system has taken a significant turn during the current party government with the introduction of a series of laws and regulations initially outlined in government plans.

The following discussion will provide more insights into the current educational decision making processes in Turkey by presenting the official discourse on education – its current purposes, objectives, strategies, and responsible actors – mainly through the higher-order policy documents; that is, government action plans, national development plans, national program for the EU Acquis, national education councils, and ministerial strategic plans respectively. Figure 12 depicts the aforementioned hierarchy in educational decision making through the influence of each of these policy documents on final decisions. This discussion will provide the foundation for deliberating the ascribed value of evaluation for educational decisions to be argued in greater detail in Chapter 5.
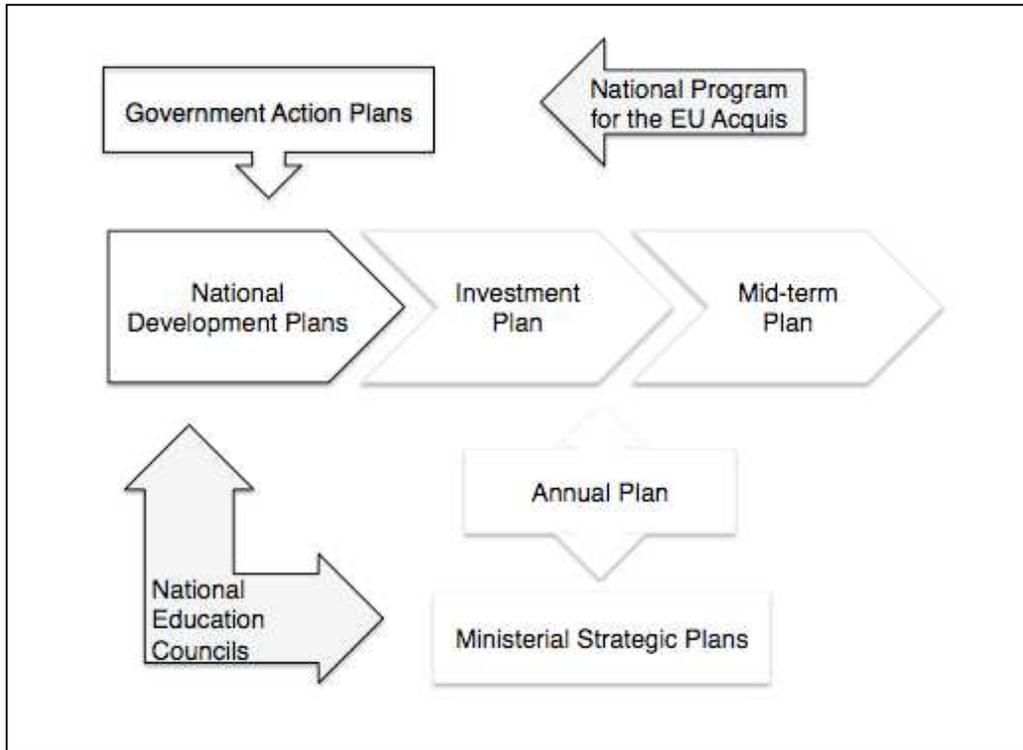
Figure 12. The Hierarchy of Official Directives in Turkey's Education Domain

Government action plans. Government Action Plans provide the primary

guidance for the design and implementation of all public programs and policies in

Turkey. Today, the need for a well-educated citizenry in Turkey is much more

pronounced in policy discussions due to rapid economic growth as reflected in the current

Justice and Development Party's governmental agenda. Indeed, the election of Justice

and Development Party (JDP, known as the AK Party) in 2002 under the leadership of

Prime Minister Recep Tayyip Erdoğan marked a new beginning in Turkish education

system. Turkey's quest to modern school of public management and education reform

has regained momentum during the JDP era, and rapid educational improvement

especially in enrollment and gender parity rates during their tenure is often cited as a

notable success story (World Bank, 2013b; Çelik & Gür, 2013; Aydagül, 2013). Political

and economic stability provided by a majority government in the parliament facilitated a

fast and sometimes radical adoption and implementation of a series of educational policies, programs, and projects geared towards increasing Turkey's competitiveness in global knowledge economy (İnal, 2012).

The JDP government signaled the upcoming changes in the education sector in their Urgent Action Plan (2003) that consisted of 250 action steps designed to achieve a more competitive market economy, economic stability, and sustainable development in the country. Twenty-one specific actions were formulized to restructure the education system to reduce historic tendencies for centralized decision making and "clumsy bureaucracy" and enable a more citizen-focused participatory planning and policymaking (Urgent Action Plan, 2003). Following the footsteps of this initial plan, the 59th, 60st, and 61nd Government Plans (2003, 2008, 2011) emphasized the quality of education as a prerequisite to realize national goals, including full EU membership, and modernize the country in accordance with the global education standards. The Prime Minister Erdoğan frequently used the terms effectiveness, efficiency, transparency, participation, and democracy as common motivators for reformulating the system. To actualize the government's commitment to eradicate every obstacle against educational attainment, the JDP government's action plans assigned the biggest number of activities and allocated the highest public spending to MoNE. The Ministry's budget allocation from the national budget in 2011 was the highest in its history, totaling almost 35 million TL (approximately US$20 million), and constituting 3.8% of GDP (Ministry of National Education, 2011).

The JDP government's pledge for a public administration reform, also providing impetus for changes in the education sector, has been materialized through a series of

legislative packages that at times faced serious opposition. The latest government plan (2011) specifically stated:

> Turkey is at a crossroads, facing a significant test. We will either claim our spot in the global system as a big, strong, and more prosperous country or we will be a country that cannot resolve its deep-rooted issues, hence continuously skid.

Motivated by a global trend in new public management featuring the principles of decentralization, participation, accountability and quality, the JDP government drafted a bill in 2004 – soon after they came to power – to reform public management in Turkey (Şişman, 2011). The former Minister of Education, Ömer Dinçer, authored a policy memorandum arguing for this reform (see Dinçer & Yılmaz, 2003). Four principles foreground the reform in this bill: decentralization, horizontal organizational structure (proposed for line ministries) performance-based inspection and control, and strategic planning. The draft bill was criticized on the grounds of destroying unitary state, which is protected by the Constitution, and dividing national sovereignty in favor of a neo-liberal, market-oriented economy (İnal & Akkaymak, 2012). Although the bill was rejected by the then President Ahmet Nejdet Sezer, bits and pieces of this agenda have been materialized through other laws and regulations (Şişman, 2011), which paved the way to significant restructuring of the education system in Turkey.

 Of these laws, the Public Financial Management and Control Law (PFMC) No. 5018 (Official Gazette, December 2003, No. 25326) is noteworthy that emphasizes result-orientation, accountability, and transparency. In a Letter of Intent to IMF in 2002 (Derviş & Serdengeçti, 2002), the previous ruling party before the JDP government provided an early rationale for a legalized fiscal control and instigated a close association between budgetary performance, and evaluation:

Over many years, Turkey's fiscal system has become increasingly fragmented, undermining transparency and aggregate fiscal control. To address this, the law will consolidate revolving funds, extra budgetary funds, and annexed and special budgets into one central government budget, under a common classification. It limits the scope for ad hoc policy initiatives, which have fiscal consequences. The bill authorizes the Ministry of Finance to set standards for accounting, financial control, and reporting throughout the general government. The law also expands the coverage of Turkish Court of Accounts (TCA) audits to all of general government and provides for external audit of TCA's own expenditures. (para. 18)

Based on this initial promise, the JDP government outlined guidelines to enact fiscal reform in public administration in their Urgent Action Plan (2003). Action Step No. 25 of Urgent Action Plan (2003) specifically called on the Court of Accounts to undertake performance-based inspection and audit, which was justified in the plan based on a global trend of goal-oriented, efficient, and effective public service delivery. The plan explicitly stated that: "This system, which has been implemented primarily in the United States and many other countries, will be applied to our country with pilot projects" (p. 30). This law now mandates that every public institution in the country develop and implement a multi-annual strategic plan that must include a clear vision and mission, measurable objectives, and specific goals to ensure efficient and effective resource allocation in accordance with national development plans. The obligation to plan strategically gave rise to the opening of Strategic Planning Departments in almost every line ministry, including MoNE, and the Department of Governance and Strategic Management at the Ministry of Development (MoD) in order to establish a culture of planning, monitoring, and goal attainment, albeit strictly associated with budgetary control.

Another striking reform on the JDP government's agenda was to restructure the Ministry as the central education authority to reduce red tape and bookkeeping so that the central administration's policy-making capacity would be enhanced. Thanks to Decree

No. 652 enacted in 2011, the number of main service units at the Ministry was reduced to 19. The order also introduced career-tracks positions to facilitate the design, implementation, and monitoring of educational policies and programs (Çelik & Gür, 2013; Aydagül, 2013).

Last but not least, one of the most controversial – and allegedly politically charged – educational laws introduced during the JDP era is the introduction of 12-year of compulsory education that divided basic education into three levels (primary, lower, and upper secondary education) of four years each, hence the popular name of "4+4+4". The Law for Eight-year Compulsory and Uninterrupted Education No.4306, also known as the Basic Education Reform, enacted in 1997 introduced 8-year compulsory schooling in Turkey in the aftermath of the military's semi-coup on February 28th, 1997. The reform initiated the abolishment of religious vocational schools, because of which, some argued, 4+4+4 was designed as a grudge-match to revive the grounding of religious premises in educational settings (see Figure 13; Okçabol, 2012). While the proponents of the legislation drew attention to the law's compatibility with educational traditions and responsiveness to the societal demands (Çelik & Gür, 2013), the opponents focused on the government's agenda to permeate a particular cultural outlook into education system most especially thanks to the reintroduction of religious vocational schools and elective courses on religion (ERI, 2013b). OECD (2013) took a stance in between noting that, "Compulsory Education for 12 years (4+4+4) can improve student transitions between educational levels, but if not managed well, it can lead to more segregation among schools and further inequities" (p. 6). A formal, official evaluation of 4+4+4 is still missing, yet think-tank organizations have already conducted empirical research on the

potential and unintended consequences of the legislation (see ERI, 2012; Çelik, Boz, Gümüş, & Taştan, 2013). The heated discussion on the pros and cons of this structural change signaled the politics of educational decision making in the country.

Age	BEFORE "4+4+4"				Stage	Grade	Age	AFTER "4+4+4"				Stage	Grade
17	General high school	Anatolian high schools (academically selective)	Vocational high schools	Imam-hatip high schools (religious)	Secondary education	12	16,5	General high school	Anatolian high schools (academically selective)	Vocational high schools	Imam-hatip high schools (religious)	Secondary education	12
16						11	15,5						11
15						10	14,5						10
14						9	13,5						9
13	Primary education schools				Primary education	8	12,5	Middle schools (different program types may apply)		Imam-hatip middle schools		Primary education	8
12						7	11,5						7
11						6	10,5						6
10						5	9,5						5
9						4	8,5	Primary schools					4
8						3	7,5						3
7						2	6,5						2
6						1	5,5						1
5	Pre-school education						5	Pre-school education					
4							4						
3							3						

Figure 13. Restructuring of Basic Education with 4+4+4 (Aydagül, 2013, p. 210).

The politics of educational policymaking was markedly clear during a recent discussion on a government bill proposing changes to the Basic Law of National Education and Decree No. 652, specifically reorganizing the structure of the Ministry – including the establishment of new divisions, the diffusion of roles and responsibilities between provincial and central administration – as well as closing down private tutoring centers, and regulating teacher and school principal appointments (Official Gazette, No. 28941, March 14 2014). While the draft bill was being prepared, and the media was stormed by nation-wide political discussions, a policy note by the Education Reform Initiative (February, 2014) remarked that the preparation of the draft bill should embrace a holistic, participatory, and evidence-based approach to regulating the education system, which did not seem to be taken seriously. One significant change is related to the role of

the Board of Education. The Board has historically been one of the highest decision-making bodies of the Ministry. According to the bill, the Board's power in designing, implementing, and modifying curriculum is to be relegated into main service divisions accordingly. The Board is now considered an advisory body only that provides research-based opinions and recommendations (related to curriculum implementation and effectiveness) to the Minister for a final judgment and decision. A reduction in the decision-making power of a historic unit at MoNE, and removing several posts from the Ministry are thought to have profound influence on the ongoing projects and programs, as well as participatory democracy and decentralized decision making (ERI, 2014).

The most controversial and political component of this proposal has evidently been the abolishment of private tutoring centers that attracted opposition across the political spectrum, and raised questions about the JDP government's agenda to contain what is known as the Hizmet movement or the Community led by a U.S.-based Turkish preacher named Fettullah Gülen – a major operator of these centers. Due to a deepening conflict of interests between the AK Party and the Community, who were once known to be allied Islamic conservatives, the government's move to abolish Gülen's main source of recruitment of loyalists contributed to the notion that the government's policies are not evidence-based per se but rather politically motivated, if anything (Şaşmaz, 2013). Despite the concerns regarding the potential negative impact of closing tutoring centers on college attendance rates and equality of opportunity for low-income students (see ERI, 2014), the draft bill was enacted, raising more questions about the government's political agenda behind the proposed changes.

National development plans. In addition to government action plans, national development plans constitute the second largest official directive – as a constitutional requirement – guiding and framing the educational agenda in Turkey for a specified period of time. The plans cover almost all sectors and industries (i.e., economy, transportation, health, education, culture, energy, welfare system, agriculture, and so on), providing "a long-term perspective and unity in objectives not only for the public sector, but also for the society" (Ministry of Development, 2006, p. 12). The Ministry of Development (MoD) (previously the State Planning Organization) coordinates the preparation and implementation of the plans to operationalize the overall vision, targets, and performance indicators for each public organization. The plans are prepared by the participation of government officials, academics, and experts from public and private sectors; approved by the Grand National Assembly, and monitored by MoD.

Similar to government action plans, the 8th (2001-2005), 9th (2007-2013), and the 10th (2014-2019) Development Plans all endeavored to secure and justify Turkey's place in a rapidly changing, globalized world where the importance of knowledge, competition, efficiency, and effectiveness is underlined. The following excerpts illustrate:

> Countries adapting themselves to the faster change in the world, equipping their individuals with the capabilities required by this new environment, having access to, producing and using information shall have an impact and will be successful in the 21st century. (Ministry of Development, 2001, p. 244)

> The vision of the Ninth Development Plan […] was determined as, "***Turkey, a country of information society, growing in stability, sharing more equitably, globally competitive and fully completed her coherence with the European Union***." (Ministry of Development, 2006, p. 13, ***emphasis*** *in original*)

> Turkey's position in the hierarchy of international cooperation and value chain is to be gradually elevated to higher levels, by mobilizing our country's potential, regional dynamics, and our people's capabilities, so as to speed up our

development in an ever changing global economy. (Ministry of Development, 2013, p. 1)

The plans consider quality education among the priority areas as a prerequisite to enhance Turkey's international competition. The motto of the 10th Development Plan (2013) is "Qualified People, Strong Society" whereby the positive association between educational achievement and economic production is strictly underlined. To achieve this overarching national goal, the plan puts a significant emphasis on education targets as evidenced by the first sector being mentioned in the report. These plans commonly associate educational outcomes with competitive labor market, setting a mission for creating a qualified workforce that produces high quality products and services demanded globally. To achieve this goal, the plans continuously set out to increase the share of public investment in education.

Nevertheless, many challenges remain against the attainment of educational targets and objectives during the specified time period that are reformulated as goal statements in subsequent plans without sufficient explanation or justification. Among these challenges are common issues related to schooling rates, classroom size, girls' education, instructional technology, and student test results. Recognizing a lack of discussion on actual goal attainment, TEPAV (2013) made the following observation in their critique of the last plan:

> Essentially, we believe that it is not assertive enough in stating the shortcomings of the earlier plans and the reasons as to why previous targets were not fulfilled. The timidity concerning the evaluations on critical structural areas (such as alleviating the informal economy, overcoming the disparity between direct and indirect taxes raising the domestic savings rate, strengthening high-value-added production etc.) weakens the determination of the report in addressing relevant reforms. (p. 11)

Their observation relates to the role of evaluation in the design and implementation of national development plans. To address the shortcomings of educational targets over time, evaluation activities and cognate fields of application (i.e., performance measurement, quality assurance, monitoring) are increasingly mentioned as tools of reference in these plans. The 8th Development Plan demanded that, "An effective monitoring and evaluation system at project level as well as national level shall be established for a prompt identification of changing conditions and bottlenecks incurred" in order to increase efficiency in planned public investments" (p. 228). Similarly, the 9th Development Plan had an explicit section on monitoring and evaluation activities for the first time in a development plan. *Implementation, Monitoring, Evaluation and Coordination* section of the Plan (pp. 113-120) envisages informing the general public about the progress in development. This section also aims to harmonize Turkey's evaluation activities with the EU Acquis.

The novelty of the 10th National Development Plan compared to the previous plans resides in its explicit acknowledgment of M&E as important mechanisms for planning and programming. M&E is closely associated with strategic planning and results-based management that are believed to strengthen accountability and effectiveness in public management and planning. Under the section of Strategic Public Management, the 10th Plan (Ministry of Development, 2013) targets the following:

> The fundamental goal is to improve effective implementation of strategic public management, and actualize the principle of accountability in all aspects of governance cycle from planning to monitoring and evaluation. To this end, the main principles are to ensure participation, transparency, and citizen satisfaction through efficient and quality public service. (p. 56)

The plan stipulates the establishment and utilization of performance management and quality assurance systems to ensure effectiveness and equity in the delivery of

educational services. The plan commissions MoD with the duty to monitor and evaluate the overall progress in achieving specified goals, yet the methods and principles of M&E are to be determined later by the Council of Ministers. Although M&E is gradually taking more physical space in national development plans, conceptual linkages between evaluation information and policy practice are yet to be discovered and operationalized.

National program for the European Union Acquis. The European Union is by far the most frequently mentioned, referenced, and admired international entity in Turkey's journey for development and modernization. Turkey's bid for the EU membership has been a significant, external force that informs and shapes national education policies and programs since the Helsinki Summit in 2005 (Esme, 2002). Compliance with the Copenhagen Criteria and harmonization with the EU Acquis (i.e., laws and regulations) gathered under 35 chapters, one of which is education, is of utmost importance in almost all policy discussions and official statements in Turkey. After the Union gained political status with the Maastricht Treaty in 1992, education has become a key area for unification, greater harmonization, and enhanced relationships between the member countries (Terzi, 2005). EU's educational policies aim at strengthening mutual understanding and cultural ties between the people of Europe; cultivating educated, competitive European citizenry; and encouraging technological innovation and development (Barkçın, 2002). The Lisbon Treaty (2000) underscores the EU's overarching goal to become the most competitive player in global knowledge economy and invites all members and candidates to align their educational programs and policies accordingly, which Turkish educational authorities endeavor to satisfy.

The National Programmes for the Adoption of the Acquis (NPAA) (Ministry of European Affairs, 2001, 2003, 2008) provide guidance to line ministries in Turkey in their efforts to converge with the EU requirements and priorities. In 2008 NPAA, Turkey is presented as a country whose fervor for modernization, and civilization can never be tamed until development is fully reflected in all aspects of life. To this end, a full membership to the EU is posed as an undisputable route to take. Education is regarded among the priority sectors to be able to compete within the EU, and the educational growth is mostly positioned within the labor market. "Qualified labor force," "entrepreneurship," and "knowledge-based competitive economy" are repeatedly mentioned and presented as among the main direction and goals for education. In 2003, NPAA outlined a very detailed action plan for harmonization in accordance with EU standards regarding compilation, organization, and management, reporting and dissemination of statistical information. This plan announced the preparation of a draft Turkish Statistical Law in accordance with Eurostat guidelines to be adopted in June 2004 (Ministry of European Affairs, 2003).

In addition to NPAA, the European Commission monitors the candidate country's commitment to implementing planned reforms in accordance with accession priorities and requirements through the annual *Progress Reports*. Although the Commission has been publishing reports on Turkey since 1998 together with regular reports on other candidate countries, the year of 2005 marked the official beginning. The reports published between 2005-2013 all have mentioned that Turkey's alignment in the area of education is almost complete – despite some setbacks in domestic politics – based on the country's progress in Europe-wide educational harmonization efforts such as Bologna

Process, the Community Programmes Leonardo da Vinci, Socrates, and Youth, Lifelong Learning, and Youth in Action (European Commission, 2006b). A discussion on monitoring and evaluation activities in these reports takes place specifically in the context of regional policies at which point the Commission frequently presents mixed achievements. In 2012, the Commission, for the first time, explicitly defined the purpose of evaluation for national policy development, and raised concerns about the lack of connection between ongoing policy development and unsystematic M&E practices in Turkish context (European Commission, 2012). The commission welcomed the launching of a web-based Integrated Management Information Systems at MoD (2006), yet they appeared concerned about the unsystematic progress in building structures and processes for monitoring and evaluating national policies. In addition, under social, employment, labor market, special education policies, a lack of impact evaluations and reliable data are considered weak points in Turkey's alignment with the EU Acquis (European Commission, 2012). Notwithstanding this concern, Turkey's commitment to improving statistical apparatus in public management has been repeatedly praised especially thanks to the passage of the Statistics Law, although the Commission remains concerned about the lack of methodology used for collecting data.

The connection between evaluation and educational policies in Turkey has been clearly made through an external project. MoNE received a 3.7 million €*Capacity Building Grant* from the European Union in 2006 to embrace new modalities of decision making so that Turkey's educational system would better harmonize with the EU policies and regulations (European Commission, 2006a). Although the results of this capacity building initiative are undocumented (ERI, 2013a), the work plan indicated that the

Ministry would undertake a series of structural changes in its planning, implementation, and monitoring capacity in order to improve educational quality and be able to compete within the Union. This is a significant statement because scholars argue that Turkey's successful membership is contingent upon quality education policies (Terzi, 2005; ERI, 2010b). After several years since the grant was awarded, one significant contribution of the proposed study is to illuminate the need and capacity for evaluation in the Turkish educational decision making context. The green paper published after this project paved the way to the preparation and the passage of Decree No. 652 as discussed before.

National education councils. Another influential force on informing educational programs and policies is the National Education Council. According to the MoNE's by-laws, the National Education Council is MoNE's highest advisory agency. This advisory body embraces a national participatory process whereby elected politicians, appointed bureaucrats, academics, civil society organizations, school principals and teachers – sometimes students – gather together to discuss the past, present, and future of education in the country, identify areas of consideration in moving forward, and propose changes and action steps. The Council does not have the legislative power; decisions are enforced if and when the Board of Education under the Ministry check their appropriateness and applicability according to educational laws and regulations, and then present the Minister for approval. Education Reform Initiative (2013a) remarked about the implications of Council's standing:

> The decisions that are made at the [National Education] Council are made with a simple majority vote by the participants without being based upon evidence. Draft decisions that are prepared during intensive commission work before the Council and the proposals put forward by Council participants go through the same process. Moreover, there are no mechanisms that will monitor how decisions by the Council are implemented by the Ministry, or in other words there are no

125

mechanisms that will increase the Ministry accountability. All these make the Council's contribution to a participatory and data-based policy-making process questionable. (p.13)

The Council's lack of legislative power ignited debate about inadequate implementation of the decisions taken by civic participation (Deniz, 2001). Aydın (1996) conducted a survey with the participating members of the 15th National Education Council (1996) with regards to their opinions about the impact of Council decisions on education policies and programs. A majority of the survey participants indicated that the influence of Council decisions on education policies is limited. Survey participants noted that the Council's place in the Ministry's hierarchy should be strengthened for decisions to have a greater weight. Yet, the Council has led to significant changes in the Turkish educational system since 1939. For example, the duration of compulsory basic education was raised to 8 years with the Law No. 4306 entered into force in 1997 (Deniz, 2001).

The Council decisions are important venues to explore the discourse around the role and utility of evaluation for educational planning and programming. A few recommendations during the 16th National Education Council (1999) clearly stated that evaluation systems need to be established and used to improve the quality and the quantity of vocational and technical training based on changing context and needs. Seven years later, the 17th National Education Council convened in 2006 with the participation of 850 elected and appointed members. Unlike the 16th Council (1999), recommendations made during the 17th Council covered a variety of issues ranging from special education to testing and examination systems. Several recommendations touched upon the importance of evidence-based practices to improve educational quality and quantity. Additionally, some of the 17th Council's (2006) recommendations specifically touched upon monitoring and assessment of educational practices. In one case, the 17th Council

recommended establishing accreditation systems to ensure educational quality in educational institutions. Compared to the 16th and 17th Councils, the references to evidence-based practices, performance monitoring and evaluation were much more limited during the 18th Council (2010). Some recommendations made an explicit case about bringing national context and values to the foreground in improving the national education system in a globalized world.

Ministerial strategic plans. Based on the expectations, goals, and objectives set out in these higher-order policy directives discussed above, the Ministry of National Education is commissioned to prepare and implement a periodic institutional strategic plan – mandated by PFMC No. 5018. As stated in Chapter 1, MoNE prepared its very first strategic plan in 2009. Similar to the discourse in previous documents, the need for strategic planning was positioned within the context of good governance, motivating an efficient balance between inputs and outputs, and enhancing accountability, transparency, participation, and performance (Ministry of National Education, 2009). Most especially, the EU Acquis and the national program for accession were frequently mentioned as main references for educational strategies in this document. The plan's novelty in launching a new educational system of governance resides in its comprehensive yet succinct description and explanation of educational goals for the entire education system.

The plan closely specifies quantitative indicators for each strategic goal under 10 thematic areas (pre-school, primary education, secondary education, vocational and technical training, private education, special education, higher education, lifelong learning, organizational improvement, inspection and guidance, and the EU Acquis and international relations) and outlines the policies and programs to be designed and

implemented to achieve each goal (see Ministry of National Education, 2009). The plan addresses the kinds of information systems to be utilized to facilitate effective decision making at the Ministry. Performance, measurement, inspection, control, total quality, research, development and evaluation are interchangeably used in the first plan. While monitoring and evaluation are conjoined regarding policy-level decisions, the terms measurement and evaluation are used together to discuss student test scores.

Activity reports published annually by the Ministry since 2011 specify the extent to which these strategic goals are indeed achieved. These reports are full with descriptive data tables, demonstrating changes in educational expenditures, teacher recruitment, or the number of schools inspected. Every performance indicator, data point, and information is quantified (see Ministry of National Education 2010, 2011, 2012). The biggest sections of the reports are dedicated to instructional technology and international education projects. The reports do not provide information on corrective actions steps for goals that were not met during the intended period. As such, the reports signal that data is collected and stored on ongoing projects and system-wide initiatives without necessarily assessing their contribution to quality of education (ERI, 2010b).

Conclusion

Motivated by her desire to be a competitive global player, education is recognized as the most important variable of national development in Turkey. In order to achieve the targeted quality, a dynamic, modern style of governance that eclectically blends global trends and Turkey's long-established history and culture has been promoted (Government Action Plan, 2003). Clearly, educational authorities put significant weight on global indicators of quality, efficiency, and effectiveness in designing, reorienting, and

legitimizing Turkey's national education agenda (World Bank, 2005). Official documents frequently refer to "global knowledge economy," "competitive market," "economic stability," "sustainable development," "qualified workforce," "participatory democracy," "rule of law," "privatization," and "modernization" as main drivers of national policies, and legitimize the need to adhere to a global trend for public administration that values accountability, transparency, and performance as main principles of governance (see Urgent Action Plan, 2003; Government Action Plan, 2003, 2008, 2011; Ministry of Development, 2013). To this end, strategic planning, performance-based budgeting, auditing, inspection, and evaluation are repeated as governmental tools for quality and performance assurance (Ministry of National Education, 2009).

The current policy discourse around international competitiveness and the global knowledge economy highlights the significance of effective policies and programs in cultivating an educated and competitive citizenry in Turkey. Especially the Ministry of National Education, the Ministry of Development, and the National Education Councils all emphasize addressing today's educational challenges with improved planning, programming, and monitoring. Although the need for improved decision making has implicitly pointed to the need for establishing and utilizing evaluation systems in the country, the value of evaluation for educational decision making from the perspectives of primary Turkish stakeholders and their official statements (policy documents, plans, and programs) has not been explored yet. The next chapter will dwell at large into this question.

Chapter V: Research Findings

This chapter provides an in-depth description of the ascribed value of evaluation in the Turkish education decision domain based on findings from individual interviews and a review of official policy documents. Findings are organized in four broad categories that emerged from data collection and analysis: (1) evolution of evaluation in the Turkish education decision domain, (2) characteristics of the educational decision-making context in Turkey; (3) Decree No. 652; and (4) positioning of evaluation within the education decision domain. In doing so, this chapter demonstrates that, while the value of evaluation has been uplifted by Turkey's quest to Westernize and modernize, and supported also by the country's development partners, the decision domain is plagued with hierarchical decision making, clumsy bureaucracy, and a lack of strategy, all together obscuring evaluation as a valued enterprise. Despite the establishment of monitoring and evaluation units at the Ministry thanks to Decree No. 652, evaluation remains as a new construct in the education domain that is closely associated with quantification, performance management, and compliance.

Evolution of Evaluation in the Education Decision Domain

This section discusses in greater detail the evolution of evaluation activities (or lack thereof) in the education decision domain in Turkey to understand the antecedents of perceptions, experiences, and attitudes towards the concept and practice of evaluation. The section shows that evaluation is a new construct in the Turkish education domain, including both governmental and nongovernmental sectors, that was originally introduced into the sector by Turkey's development partners. Despite the notion that Turkish culture embraces traditions of critical thinking and justice that are compatible constructs with

evaluative thinking, perceptions about evaluation are largely affected by the way Turkey's aid partners introduced the concept. To develop this argument, the section will discuss the evolution of evaluation in governmental and nongovernmental activities and how this evolutionary process has been facilitated or interrupted by international and cultural processes.

Governmental activities. Overwhelmingly, informants mentioned the short yet emerging history of taking evaluation seriously in the Turkish education decision domain and argued that evaluation has hardly, if ever, been a part of the fabric of educational decision making. They indicated that there have been neither systematic mechanisms nor an established mentality to necessitate and conduct formal policy or program evaluations due to a perceived lack of awareness of the value of evaluation for governmental decisions. One official reported that the National Development Plans repeat the same goals and indicators year after year; they are basically "copied and pasted" from the previous years without acknowledging what has been achieved over the course of years. The official remarked: "Monitoring and evaluation in our public management is not professional, but only an imposter. No one asks about the achievement of goals in previously approved plans or questions the repetition of the same goal in the subsequent plan." Another government official agreed with this sentiment and added, "Those who think they were conducting evaluations had not gone beyond counting the number of schools." One civil society representative summarized the lack of history with evaluation as follows: "Monitoring and evaluation is the Achilles' heel in the Turkish education decision domain."

A few informants disagreed, claiming that Turkey possesses a long-standing administrative culture so it would be naïve and "unjust" to claim that evaluation has never been part of its educational system. One government official complained: "So we would consider an activity evaluative only when we use the word *evaluation* for it?" A few informants referred to personnel investigations, internal control mechanisms, national development plans, budget plans, strategic plans, and academic research as evaluative activities, strengthening accountability and legitimacy in the educational domain. One professor of education noted that the first participatory evaluation was indeed conducted at *Village Institutes*, where teachers and school administrators would gather regularly to provide constructive feedback on teaching and learning, as described in Tonguç's (1947) *Canlandırılacak Köy* [*Village to Be Revitalized*]. According to these informants, the fact that ongoing activities were not specifically called evaluation does not mean that evaluative activities were never done in Turkish education.

To support this latter argument, a few other informants mentioned additional previous attempts to obtain and utilize research evidence for educational decisions. First, prior to Decree No.652, EARGED (Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı) [Department of Educational Research and Improvement] and APK (Araştırma, Planlama ve Koordinasyon) [Research, Planning, and Coordination] conducted action research for ongoing educational initiatives. EARGED was established as a part of World Bank's National Education Development Project (1990) with a loan of US $90.2 million to regularly and systematically conduct and utilize scientific educational research to improve educational effectiveness (see EARGED 1997, 1999, 2002, 2005, 2009, 2010). APK was established to build the planning capacity of the Ministry, which is now the

duty of the General Directorate of Strategy Development. These units were closed with the passage of Decree No. 652 to reduce overlap in responsibilities and functions across units. Second, the Court of Cessations and TÜBİTAK (Scientific and Technological Research Council of Turkey) were mentioned by a couple of academics because of their work monitoring educational expenditures, although these informants also indicated that evaluation is much more nuanced than mere auditing. Despite these existing or discontinued systems of evaluation, a substantial consensus emerged in the interviews that "[w]e have been stubbing toes and tripping, [but] we never could obtain the level [of evaluation] we desired."

Nongovernmental activities. As in the governmental arena, a majority of the informants argued that evaluation is an emerging concept and practice in nongovernmental arenas in Turkey. Government officials and academics alike critiqued the state of academic work on educational research and evaluation in the country. Some argued that this limited academic work inhibits the visibility of evaluation and reduces the chances of highlighting its value. One government official indicated that, when they were preparing a monitoring and evaluation (M&E) workshop, they could not locate one single academic research, report, or working paper on evaluation terminology, methodology, and processes specifically in the Turkish context. They instead utilized international resources authored by the World Bank and OECD. Another government official echoed this point, arguing that Turkish academics often conduct research on the finance or economics of education, not necessarily policy planning, processes, and outcomes.

Lack of issue focus was also evident in the informants' background. Very few academics interviewed for this study reported research interests focused on educational policies and interventions. Instead, most conduct basic research at the micro-level – in classrooms or schools. Only two graduate-level courses at two different research universities on educational policymaking and evaluation were mentioned during the interviews.

While Turkish higher education institutions pay very little attention to evaluation, civil society organizations have assumed a major role in elevating the role of evaluation in educational decision making. Government officials have partnered with civil society organizations and in general seemed more content with the quality of their studies than with academics'. Officials often mentioned the Education Reform Initiative (ERI), SETA (Foundation for Political, Economic and Social Research), as well as TEDMEM (Turkish Education Association) and TÖDER (Association for Private Tutoring Centers) during interviews. Among these actors, ERI's targeted focus on evidence-based decision making merits attention. Their *Education Monitoring Reports*, published annually, explicitly address the need for an evaluation agenda in the education domain. The following excerpts from their reports illustrate this observation:

> One of the most important long-term steps that can be taken for the education system in Turkey is to reconstruct policy-making processes in education on an evidence-based, participatory and transparent basis. (ERI, 2010b, p. 6)

> In accordance with one of the key principles of good governance, *transparency*, MoNE [Ministry of National Education] should continue to share data on student absenteeism, class repetition, and dropouts. We strongly believe that this would lead to independent, objective, evidence-based evaluations of Turkey's education system, paving the way for improvement. (ERI, 2013b, p. 8)

In sum, given the limited history with the concept and practice of evaluation in Turkey, civil society organizations' role in highlighting the value of evaluation for policymaking,

hence filling in the academic gap, was praised during by the educational authorities interviewed for this study.

Turkey's development agenda. The repeated remarks about Turkey's short history with the concept and practice of evaluation uncovered the term's trajectory in the education domain. Informants often mentioned that Turkey's quest to modernize and westernize has resulted in it embracing many educational concepts, ideas, and processes from other countries, including evaluation. These concepts often enter the Turkish decision domain through many projects and programs funded by international donors such as UNICEF, World Bank, and the European Union (EU) (see Table 15).

Table 15. A Selection of International Projects Implemented in Turkey

Major Funder	The Name of Project
UNICEF	• Pre-School Education Project (2010-2013) • The Childhood Development and Education Project (2011) • Child-Friendly School Environment Program • The Catch-up Education Program (2008-2009) • The Girls to Schools Now Campaign (2001-2005)
EU	• Democratic Citizenship and Human Rights Education in Turkey (2011-2014) • Improving the Quality of Vocational Education and Training in Turkey (2012-2014) • Strengthening Special Education Project (2011-2013) • The Information System for Determining Educational Needs on Vocational and Technical Education Project (2005-2013) • Strengthening the Statistical Capacity of the Ministry of National Education (2010-2012) • The Strengthening Special Education Project (2011) • Strengthening of Vocational Education and Training Programme (SVET / MEGEP) (2008-2010) • The Vocational Education Project for Employment (IMEP) (2009) • Capacity Building Support for the Ministry of National Education (2008-2010) • Support to Basic Education Project (SBEP) (2002-2007) • The Modernization of Vocational Education and Training Programme (MVET) (2003-2006)

Major Funder	The Name of Project
World Bank	• Fatih Project (Movement of Enhancing Opportunities and Improving Technology Project) (ongoing) • Conditional Cash Transfer Program (2003-ongoing) • The Secondary Education Project (2006-2011) • The Basic Education 2 Project (2002-2005) • The Basic Education 1 Project (1997)

One academic claimed that M&E entered the Turkish educational domain with the EU's 2008-2010 project on "Capacity Building Support for the Ministry of National Education," one component of which focused on strengthening evaluation systems at the Ministry. Another project that is worth mentioning, as far as the rise and fall of evaluation in the Turkish education domain is concerned, is *Monitoring and Evaluation for Development Programs* (MEDP) implemented by the Ministry of Development (MoD) and supported by the World Bank during 2007-2010 (European Commission, 2010). The project was initiated to strengthen and incorporate elements of results-based management (IRBM) and citizen responsiveness into the 10th National Development Plan so that national decisions will be improved based on M&E data (Stout, 2010). As part of project activities, World Bank consultants provided a workshop – *M&E for Policy Formulation and as an Integral Part of Government Institutions* – to 20 bureaucrats at MoD and pilot institutions (Ilgın, 2010). In addition, World Bank's *10 Steps to a Results Based Monitoring and Evaluation* by Kusek and Rist (2004) was translated into Turkish (Ilgın, 2010).

While a few informants argued that donor funded education projects meet the essential need for research and development (R&D) in educational planning in Turkey, some (N= 10) problematized the excessive borrowing from other education systems and

called it "the disease of conceptual borrowing." According to these critiques, evaluation

is inherently an Anglo-Saxon concept and tool that Turkey appropriated in her efforts to

be a global player starting around the 1980s. These informants believed that evaluation

and its cognate terms – total quality management, performance management,

accountability, transparency, strategic planning, teacher credentials, and school standards

– denote a neoliberal market agenda for educational planning and programming modeled

after the field of business administration in the United States. Specifically, this neoliberal

market agenda is believed to conceptualize evaluation practice as a tool to demonstrate

the value chain of an organization – public or private – through a production function

model, which facilitates the commodification and quantification of educational processes

and outcomes. One academic argued that evaluation's neoliberal foundations indeed feed

a culture of competition, which is found to be at odds with Turkish way of life, and

continued:

> Western knowledge is a new form of hegemony. We try to understand and get to
> know ourselves through their methodological lens... I think the Western concepts
> tend to control us. They create mechanisms through which I could surrender my
> will to the command of free market.

These respondents believed that the concept of "evaluation" was borrowed

without understanding the context within which it was developed; thus, its value for

Turkish educational decision making has been justifiably limited. They believe that every

concept is the byproduct of its unique context addressing distinct problems; as a result, it

may not be compatible with the local realities of a new setting. The following comments

illustrate this feeling:

> Unfortunately, we are borrowing educational concepts without positioning them
> in their unique context and understanding under what circumstances and why
> these concepts have been created. We accept them literally as if they represent the

universal truth. This occurs due to a failure to link these foreign terms with the current societal ties and realities.

It is like you are uprooting a tree from its original soil and planting it in your garden with the hope that it will bush out and blossom. It is the same in science. You need to take the entire concept with its roots. You cannot succeed by trimming down or whisking off.

Due to a perceived lack of contextualization of evaluation, three academics and two officials indicated that there is a tendency in Turkey to reduce the entire field of evaluation practice to the level of methodology or, more specifically, statistics. One academic called this tendency "methodological fetishism" whereby existing evaluation or cognate activities embrace a methods-first approach, which is believed to perpetuate the Western bias in educational programs and policies without accommodating contextual nuances. Thus, these informants felt that the borrowed concepts' credibility, relevance, and usability in Turkey should be tested in the context of Turkey's historical, cultural, and political realities.

Turkish culture. Due to evaluation's perceived foreign origins, a deeper understanding of Turkish cultural context in facilitating or inhibiting the value of evaluation appeared relevant in a number of interviews. According to half of the academics interviewed, Turkish culture houses a rich blend of contexts and histories spanning continents and centuries – the Hun, Gokturk, and Seljuk states and the Ottoman Empire, to name a few – and has thus gathered, adapted, and internalized a plethora of diverse concepts, processes, systems, and mechanisms, including critical thinking and judgment. Specifically, the traditions of critical thinking and justice were mentioned as important to building evaluation's value.

Some academics specifically mentioned the value of constructive criticism inherent in Turkish culture as a necessary tool to create an archetype of an "impactful

human" who learns from his/her mistakes. According to a few academics, this critical aspect of Turkish thinking is evident in the literary critiques from the eras of Tanzimat (Reform Movement) and Servet-i Fü'nun (Wealth of Sciences) in the Ottoman Empire from the late 19th and early 20th centuries. This culture of criticism is believed to align with evaluation as it encourages modifying one's actions based on critical feedback.

A few informants referred to a sense of justice inherent in Turkish culture that was later strengthened by Islamic traditions. They believe that Islam promotes "rightful due" and pursuit of equity in all endeavors. The understanding of "What have you done to love the God (Allah) today?" was quoted twice to illustrate the religion's encouragement of just behavior. These informants acknowledged that Western countries systematized and routinized critical thinking by creating structural methodologies, which by no means indicates that Turkish culture inherently lacks evaluative thinking in its character.

While Turkish culture is believed to be highly compatible with evaluation culture, a few informants still problematized the logic of evaluation and its inherent connection to Western culture. More than half of the academics interviewed (N = 8) argued that Western culture is based on the premises of individualism, competition, materialism, rationality, and production and consumption that organically necessitate and legitimize the need to measure, assess, and evaluate efforts to provide efficiency and effectiveness. On the other hand, Turkish culture, representing the larger Eastern mindset, promotes collectivism, collaboration, empathy, harmony, and the blending of both material and spiritual constructs, all of which may not resonate with an impulse to rationally measure and assess every human activity. One academic summarized this notion as:

139

I have never annulled the contract of a research assistant in my department so far. I indeed would have to terminate half of the contracts by now according to global administrative principles and criteria. But I simply don't. There is only one reason: humanistic concerns. They either just got married or something else. Here is what I am trying to say. The models we are trying to borrow from the West are based on rationality, and rationality does not work on its own here. It is rationality plus soul. You just have to add that component. Rationality has helped the West prosper; it is just not working here.

Given the official discourse on modernization and development, as discussed in Chapter 4, these critics suggest that Turkey might have superficially adopted Western discourse around results-based management or global governance, thereby suggesting that she also embraces the Western concept of evaluation as a way to become more of a global player, but these concepts and evaluation may mesh poorly with Turkish culture.

However, these informants cautioned that evaluation's link to the Western world does not suggest it should be abandoned altogether, but rather should be embraced in a manner consistent with contextual realities and the wealth of options available through Turkish culture. Most of the informants indicated that it is of utmost importance for future educational leaders in Turkey to know and understand their own cultural and historical context and recognize their own culturally-embedded values and norms to guide the development of contextual solutions to their own problems.

Characteristics of Turkey's Educational Decision Domain

Given the antecedents of the perceptions and experiences related to the concept and practice of evaluation, this section will explicate the defining characteristics of the education decision domain in Turkey that either enable or disable building the value of evaluation for decision making. Findings suggest at least four dimensions of the domain that matter for evaluation's value: (1) the authorities' decision-making style, (2) the

politics of educational policy, (3) national education strategy, and (4) bureaucratic

culture.

Figure 14 depicts the relations among these characteristics. In Turkey,

centralization and politicization of educational decision making in the hands of a few

political elite is believed to impede the articulation of a unifying national education

strategy and in return assigns an *ad hoc* value to evaluation. In the meantime, a perceived

lack of education strategy in the country is exacerbated by a perceived lack of planning

capacity and a clumsy bureaucratic culture that further limit the opportunities for

analytic, intellectual activities, including evaluation, to be utilized during policymaking.
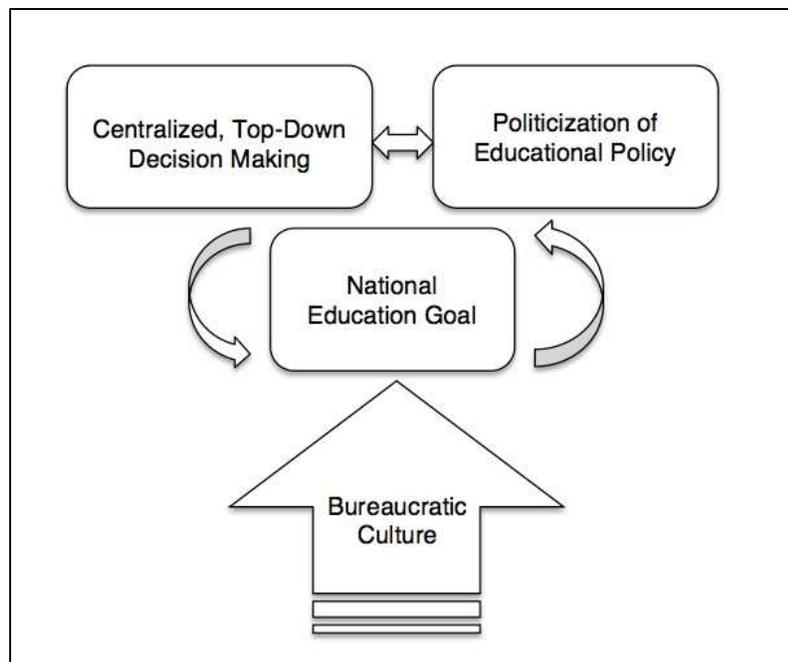
Figure 14. Characteristics of Turkey's Educational Decision-Making Context

The decision-making style. The decision-making style of governmental

authorities was the most frequently discussed characteristic of the education domain

during interviews. According to most informants, educational decision making in Turkey

is at the mercy of top administration. In a system where certain people gate-keep the decision domain, evaluation is thought to be limited to providing the information that decision makers want to hear, which in turn makes evaluation's value and legitimacy as a decision-making tool questionable. One academic associated this self-reinforcing relationship between hierarchical decisions and perfunctory evaluation to the Evil Queen's talk with her magic mirror in the story of Snow White: "Magic mirror in my hand? Who is the fairest in my land?" The mirror answers, "My Queen, you are the fairest in my land."

Specifically, academics expressed their frustration with the highly centralized and hierarchical decision-making process whereby the ruling government party is believed to dictate the desired course of action without consulting with major stakeholders about policies' implications or long-term consequences. Thus, they likened the education domain to an "elitist" platform for "social engineering." One academic believed that, when individuals oppose the educational authority, they are ostracized and alienated, which in turn encourages a culture of disengagement from governmental decision making and eventually passive obedience in public. Academics also problematized the nature and quality of the participation of nongovernmental stakeholders when they are invited to advise decision making. One academic specifically critiqued educational officials' desire to involve like-minded stakeholders in educational planning and policymaking and noted, "Incompetent leaders gather incompetent people around them."

Government officials tended to confirm this sentiment by defining a clear line of authority over educational decisions. According to them, educational policymaking follows a centralized path starting with government plans issued by the current

government, followed by National Development Plans (NDP) prepared by the Ministry of Development (MoD) and Strategic Plans developed by the Ministry of National Education (MoNE), which was discussed in Chapter 4. Within MoNE, the Minister, the Undersecretary, managers of each directorate, and branch directors are believed to inform policy decisions, in turn.

Despite Decree No. 652, designed to empower bureaucrats and enable the inclusion of other stakeholders – as discussed in depth later – bureaucrats still view themselves as policy implementers, not policymakers. A government official was taken aback when asked about experiences regarding educational planning, and remarked:

> Your request assumes we are closer to decision making; we are actually implementers rather than decision makers. Decisions are made at the top such as at the level of general directorate, undersecretariat, and minister. We only fill in the bottom by implementing these decisions and providing alternatives, and recommendations based on our research. We don't have the authority to make decisions on our own.

Themselves discontent with the level of inclusiveness in decision making, education officials noted that they invite stakeholder participation as much as they can. For example, they reported their ongoing willingness to partner with civil society organizations and/or academics on educational projects. While academics blame educational authorities for disregarding stakeholder viewpoints, government informants expressed their disappointment with academics' apathy and unresponsiveness to requests for proposals.

All in all, more than half of the officials interviewed confirmed that the Turkish education system is extremely large, multi-stakeholder and multi-geographic, and cutting across 81 cities and 892 provinces. This is thought to necessitate central decision making, coordination, and inspection, which MoNE "cannot easily let go."

Unlike academics, civil society representatives appeared more reconciliatory and argued that the policymaking process, underneath, resembles a laboratory wherein different ideas are exchanged, debated, and tested; some are discarded, and some get put on the national education agenda. Thus, it would be naïve to expect that any one single individual or organization could change the course of policymaking in Turkey.

Two representatives also believed that the policy dialogue in Turkey was minimally inclusive before the 2000s, when the election of the Justice and Development Party (known as the JDP government) opened a window of opportunity for a more inclusive policy dialogue driven by their "pursuit of legitimacy." The JDP government was believed to "normalize politics" by eliminating military and judiciary patronage, which in return cultivated a culture of responsiveness to public demand and societal needs. While one participant noted that "public policy used to be under government monopoly," they also noted that the pieces of the puzzle of participatory democracy are falling into place gradually in Turkey, which will further feed the demand for governmental responsiveness.

Nevertheless, four respondents thought that centralized, hierarchical decision making does not necessarily impede evaluation. One academic indicated that a central authority could quickly elevate the significance of evaluation without receiving so much opposition, whereas a more democratic and participatory system would slow the process of adoption and adaptation. Another academic proposed, "There is advantage to authoritarian systems [where] authoritarian bureaucracy is the sole power, the center. If the Minister is convinced and willing, impact evaluations can be done easily." Another informant confirmed this point: "If the person at top of the system in the ivory tower has

interest in evaluation, it only takes a single, signed document [mandate] to initiate the process" of streamlining evaluation across the educational decision domain.

Politics of (educational) policy. According to nongovernmental informants, Turkey's centralized, hierarchical educational decision making enables the prevalent role of politics and ideology in educational planning and policymaking. Most informants discussed the historic role of education in Turkish nation building. A couple of them quoted John Dewey, who claimed that the main role of the Turkish education system was to establish, strengthen, and transfer the national identity embraced by the new republic (see Boydston, 1983). Informants from academia firmly believed that education is a political battleground in Turkey that each government utilizes as an ideological tool. They also posited that this extreme politicization raises concerns about the lack of credibility of any available evaluation finding, since the strong political influences in educational policymaking dictate that any endeavor to obtain scientific evidence (such as via evaluations) is inherently political because evidence is obtained to legitimize the change in the political landscape.

Changing educational administration was identified as both the greatest cause and an effect of politicization of educational policies, gradually obscuring the value of evaluation in the decision domain. First, a majority of informants felt that educational policymaking in Turkey is a "grudge-match," "almost like revenge" such that every government attempts to uproot the previous government's agenda. One government official described this attitude as: "After me, the flood" [*Après moi le déluge*]. One civil society representative argued that when educational officials believe that they will not stay in the office long enough to bear responsibility for their endeavors, they engage in

short-term planning, making education all the more vulnerable to personal whims and political contestations and less susceptible to the impact of evaluation.

Half of the government officials interviewed specifically complained about the changes in top administration that they believe change the educational agenda and reset the institutional memory at MoNE. They referred to the fact that MoNE has had four different ministers within the past decade who served from 5 months to 6 years and argued that, even though these ministers were serving the same political party (i.e., JDP) each brought his or her own agenda and cadre of staff. Five officials claimed that new leaders cannot only shelve the previous ideas, projects, and policies, but they can also dramatically refocus public attention on new alternatives. Thus, these informants consistently described the education domain as a "jigsaw puzzle" whereby educational targets and objectives are constantly shifting from one government to another, enabling political interests to be the governing paradigm. As a result, they argued that abrupt changes in educational administration relegate the evaluation function to an *ad hoc*, unsystematic role in decision making and reinforce the fluctuating emphasis on evaluation as a valued enterprise.

Unlike academics, who seemed eager to problematize the politics of educational policies, civil society informants accepted the political nature of policymaking, as they believe policymaking is not a zero-sum game between politics and scientific evidence; there is indeed enough room for both, and it is naïve to demand the removal of politics from policy. These respondents claimed that everyone has a political stance, including their own organizations, the question is whether people and organizations make their political stance transparent. To this end, they work to expand the space within which

evidence can play a larger role in policymaking through respectful and responsive dialogue. In sum, similar to academics, civil society informants believed that political ideologies play a significant role in policymaking, yet they appeared more realistic about their influence and the uses of evaluation they can accommodate.

National education strategy. Informants' comments revealed that the treatment of education as a political tool precludes the development of a unifying, national education strategy for all and, in return, the development of evaluation as a valued enterprise in the decision domain. Four nongovernmental respondents who demanded answers to the following questions illustrated the lack of a national education strategy in the country:

- What is the foundation of the Turkish educational system: democracy, justice, ethics, and/or relationships?

- Who is really the subject of the Turkish education system: students, teachers, leaders, politicians, and/or foreign partners?

- What values do we want future generations to inherit?

- What kinds of leaders do we want to raise?

Informants suggested that a clearer sense of direction in policy is necessary for evaluation to offer a valuable contribution to educational decision making in Turkey. They argued that evaluation might not find its niche in an environment where educational goals and strategies are not clearly articulated and often shift.

The mutually reinforcing relationship between a lack of unified approach to national education and a lack of feedback to sustain this haphazard approach was explored by Gür and Çelik (2009) from the Foundation for Political, Economic, and Social Research (SETA), one of the most influential policy research institutes in the country. They argued

that the Turkish education system does not represent a system at all, due to the mismatch between educational goals, processes, and lessons learned, including evaluation. They posit that a lack of national education strategy leads educational authorities to "save the day," making sweeping decisions, without their decisions being grounded in systematic feedback. Their observation was explicitly shared by one of the academics who compared the current situation to "schooling behavior in fish." The professor used the metaphor rather loosely to argue that educational authorities do not follow a coordinated, targeted pattern, instead altering the course abruptly depending on the political climate. Instead, the education domain is often conceived as a "project dumpster" – one of the most commonly used expressions among all participants – of isolated, free-flowing initiatives that do not serve a common, valued purpose.

According to more than half of the academics, one factor contributing to a lack of national education strategy is a lack of planning capacity. This is largely attributed to the lack of human capital in the education domain. These informants often mentioned educational authorities' lack of background and expertise in the field of education. Instead, current recruitment policy seems to favor individuals with backgrounds in economics, management, or even engineering instead of educational policy and administration. These informants believed that this reinforces the development of narrow-minded, shortsighted educational policies and programs focused primarily on quantity of education, leaving quality of education a hit-or-miss target. Since these authorities lack proper expertise, political interests easily penetrate the educational domain, as previously discussed. As one academic remarked, "Incompetent people hide behind ideology." In the absence of long-term, holistic planning, policy ideas are randomly selected and enacted,

which, according to one official, contributes to the educational domain's image of "a random generator of educational initiatives."

In the absence of a national education strategy, a number of respondents felt that the Turkish education system borrows excessively from Western – sometimes referred to as Anglo-Saxon – countries; educational concepts, processes, and methods are randomly selected and implemented to remedy problems, and success is judged based on Western criteria. As a result, a majority of nongovernmental respondents requested that educational authorities re-conceptualize the national education strategy through an inclusive process that reflects the societal, cultural, and historic realities of the Turkish people. Excessive educational policy borrowing in return is believed to destroy the sense of ownership and commitment to those policies, resulting in less sense of valuing of evaluation. One education professor summarized the role of evaluation in the absence of an original national approach to policymaking:

> This country has not produced anything original for over 200 years. People would like to monitor and inspect things that they created themselves. When you borrow something or you are not a part of its development, then you would not have the desire to monitor.

In sum, accompanying political factors such as unsustainability of and abrupt changes in education policies and reform agendas (as discussed above) and an excessive borrowing from other countries reduce the ownership of the policies. This perceived lack of ownership disables the development of a comprehensive approach to education and further obscures the value of evaluation.

Bureaucratic culture. Academics, government officials, and civil society representatives alike frequently used terms such as "traditional," "hierarchical," and "clumsy" to describe Turkish bureaucratic culture as a historic barrier to effective

policymaking and accountability. They claimed that Turkish bureaucratic culture was largely inherited from the Ottoman Empire, which is one of the "oldest, entrenched, old-boy networks" with strong, deep-seated tendencies that are hard to overturn. One official stated, "Customs lie at the heart of our bureaucracy. These customs as old as 400 years dictate that one leader speaks, and we accept and follow suit. This is the only way it goes, no matter what they say."

One of the deep reflexes of Turkish bureaucracy identified by the respondents is a discouragement of analytic thinking exacerbated by a perceived fear of criticism among educational officials. Nongovernmental respondents believed that officials take criticism as "personal offense" or "political critique." The following remarks illustrate this sentiment:

> Let's say the Ministry prepared a 20-page document on an educational issue or concern including all data, graphics, etc. It looks just neat but they cannot publish it unlike in England, Finland, or the U.S. Why? Because they are afraid that [political] parties will issue a parliamentary question or a newspaper will critique their work in an op-ed article.

> I believe [our bureaucratic culture] is the extension of Ottoman bureaucracy from the 1700s and 1800s that was modeled after the French example. This has ingrained into our souls so much so that I even asked you if you have official permission to interview me. I would have told these things to someone else anyway. There would not be any difference in my responses. There is no secret. But we have internalized such a reflex. It is so hard to step out of this frame now. It is a strong wall to demolish.

In its aversion to criticism, bureaucratic culture is believed to prevent the development of a non-threatening intellectual platform within which new ideas can be bounced around, debated, critiqued, implemented, and evaluated. Nor does it enable appointed officials the freedom to expand their horizons and create new ideas. One academic who once served as a high-ranking official at the Ministry remarked: "It is so hard to be intellectually

productive in our governmental system. There are two reasons for it: First, the bureaucratic system is inhibiting creativity. Second, you cannot be intellectually productive with the civil servant mentality." Since rank-and-file bureaucrats have job security, they are believed to shy away from critiquing the system. This in return is thought to aggravate the authoritative nature of bureaucratic culture such that it is business as usual that no one dares question; it remains intact as a result.

Another factor contributing to the bureaucratic clumsiness identified by the informants is excessive, routine red tape and book keeping that reduce the level of quality engagement with actual educational planning and policymaking. One academic who regularly works with the educational authorities argued that educational plans remain rhetorical mostly due to massive amounts of required paperwork and official bookkeeping. One official mentioned the difficulties in enticing school principals to provide input for decision making. The official complained that:

> Workload is so heavy that school principals do not have the time to intellectually engage in research. They could suddenly come up against a brick wall by just receiving an official document that they have to deal with for about a week but nothing else. This is what we deal with right now.

Another characteristic of Turkey's bureaucratic culture is the perceived lack of communication and collaboration across different units within the Ministry. This is believed to reduce the odds of coherent educational planning and effective results. A couple of government informants shared anecdotes about competition between project offices for funding. One official cautioned that individual units may narrow-mindedly "sabotage" alternative courses of action, without providing further details. This further contributes to the "project dumpster" image of the education domain in that units work in isolation on different projects that are not integrated into the entirety of the system.

In sum, many informants shared the same sentiment: Clumsy and excessive bureaucracy does not facilitate utilization of intellectual property. This rampant limitation feeds what one government official called "institutional bigotry," whereby educational authorities consider every critique malevolent and politically charged. Given this culture, one official demanded that evaluation studies should be careful to avoid being themselves convoluted and irrelevant, lest they add to this clumsy bureaucratic culture. Thus, the value of evaluation is believed to depend on the extent to which it is providing solid, practical, tangible solutions to educational problems at hand. There is a perceived risk that evaluation can easily become a part of routine red tape by becoming so overwhelmed with the bureaucratic process that it may not yield any timely findings and thus may lose its utility.

Decree No. 652: "A paradigmatic, radical change"

Despite the common notion among the study informants that evaluation remains in limbo in the Turkish educational context, a window of opportunity to highlight its value and elevate its role in the decision domain opened up with the passage of Statuary Decree No. 652 in 2011. Not only did this decree radically restructure the Ministry of Education, but it also introduced new terminology to the public administration of Turkish education (i.e., strategic planning, performance budgeting, evaluation), leading to more intense theoretical, practical, and political debates about the Decree's long-term implications. The Decree reiterates that it is the main role of the Ministry to design, implement, inspect, monitor, and evaluate national education policies and strategies for every level of education. In doing so, the Articles define the specific roles and

responsibilities of each service unit by highlighting their role in policymaking and evaluating those policies.

The informants' remarks revealed that Decree No. 652 is already a landmark in the Turkish educational domain in its attempt to alleviate if not fully eradicate the problems perpetuated by a centuries-old legal and institutional framework as previously described by the informants. This law was consistently cited as "strong," "essential," and "inevitable," and it is thought to represent "a paradigmatic, radical change" in Turkish education history. Specifically, it has paved the way to systemic changes to Turkish education: 12 years of compulsory education, the creation of elective courses (i.e., native language and religious courses), and the elimination of the course on National Security and the directive on school uniforms. As such, according to government officials, this decree set in motion the most desired and needed changes in the education sector as discussed for over three decades in almost every major national and international development report.

Decree No. 652 essentially targets the capacity of the Ministry to enable educational authorities to produce effective policies. Ömer Dinçer, the mastermind of the Decree, stated, "I have never come across an institution like MoNE before where there is tremendous amount of data but little use of it due to non-coordination across units." One academic called the previous organizational structure "irrational" and commented:

> How can you have 50 different units at the Ministry that deal with exactly the same topic area? … All of these were indeed overlapping. Communication and coordination were broken among these units. You could see this point in World Bank's reports on Turkish national education.

Another official described the old system as "historic tribal communities at the Grand Canyon back in the day where they were scattered all around and had not known about each other's existence for decades."

Hence, the directive envisioned an education system that connects the components of structure, goals, human capital, budget, culture, and technology whose interconnection is believed to be missing in the previous institutional structure. For example, to increase effectiveness, responsiveness, and flexibility in decision making, Decree No. 652 created a matrix management system in which similar tasks and skillsets are brought under the same work unit (see Table 6 and 7 in Chapter 3 for the new MoNE organizational framework). In sum, the goal was to break down silos of function by assigning specific roles to each unit so that effective policies are designed and implemented based on ongoing evaluation information.

Overwhelmingly and consistently, Dr. Ömer Dinçer, the former Minister (2011-2013), was cited as the main proponent of Decree No. 652 and a vital figure to initiating an evaluation regime in the Turkish educational domain. His championing of evaluation was considered unprecedented. Mr. Dinçer was regarded as a powerful actor with a strong background in public administration and management. He has also been a published professor of management who is known to value academic research and scientific evidence. One civil society representative indicated "[t]he appointment of Dinçer as the Minister was a strong indication that the JDP government [once] placed high significance on education." An academic described him as follows:

> He had his own rationale, a sense of activism. Just think about it. He is the man who changed the playing field in the 8th year of a ruling party – which has been in power for 10 years – in a field that mostly his fellow party men embodied. You have to be really bold to be able to pull off something like that because you need

to defy the long-standing web of nepotistic ties. Dinçer was just a direct, honest man. Thus, he could do it; he was the only one who could do it.

In a decision domain where evaluation is considered the Achilles' heel, almost every official interviewed for this study praised Mr. Dinçer's enthusiasm for putting evaluation on the radar during his tenure. Many informants implied that if Dinçer had not been an intensely results-oriented person with a strong academic background, it would be hard to imagine the passage of Decree No. 652, hence the creation of the M&E units and the possibility of contemplating the value of evaluation in educational decision making. These impressions signal the influential role individuals can play in fostering the value of evaluation, in addition to legal framework.

Legal mandate for evaluation. One intriguing subtle recurring theme in government officials' discourse around Decree No. 652 was the power of a legal framework in identifying, regulating, and guiding norms of behavior and setting criteria for judging that behavior in the education domain. A majority of the officials interviewed shared a sentiment that once an evaluation mentality is internalized, the possibilities of evaluation are infinite because of the legal enabling environment.

The legal environment for gradually building the value of evaluation has been enabled over time by two official strategic acts, one concerning the entirety of government and the other addressing specifically the organization of MoNE. First, PFMC Act No. 5018 introduced strategic planning and ongoing M&E activities, albeit strictly for budgetary purposes, into Turkish governmental life. This act was originally formulated under the Fiscal Structural Adjustments of IMF Stand-By Agreements. The PFMC reflected the JDP government's desire to reform and restructure public administration in accordance with standards for good governance prescribed by

international partners, setting in motion a series of laws and regulations that gradually embedded an evaluation mentality into Turkish administration. A few informants confirmed that Act No. 5018 introduced several new concepts into the educational system such as accountability, transparency, internal control, and inspection.

In Act No. 5018, evaluation is disguised under the umbrella term of "strategic planning"—a new construct in Turkish public administration. A 2006 internal survey of 134 senior officers at MoNE concluded that officials did not have enough knowledge about the concept and process of strategic planning, yet they believed that strategic planning would improve institutional learning and management, contributing to the improvement of educational policies and programs (Türk, Yalçın, & Ünsal, 2006).

Second, under PFMC leadership, Decree No. 652 was enacted. As described above, it officially institutionalized the evaluation function within MoNE, highlighting and mandating the use of evaluation as a decision-making tool. It was not until Decree No. 652 that evaluation was specifically named and discussed as an integral part of educational policymaking. Government officials often mentioned the legally binding character of evaluation, simply meaning they are "required to do it now," meaning conducting evaluations; otherwise, violations are penalized. A few informants confirmed that before Decree No. 652, in the absence of a legal mandate, evaluation was dependent upon individual discretion. Because of this mandate, they now seem not to question the value of evaluation.

Despite a legal environment embracing evaluation as a decision-making tool, half of the government officials interviewed as well as some of the academics believed that the institutionalization of evaluation through a legal mandate does not necessarily

translate into sustained value of evaluation. In other words, a legal mandate to conduct evaluations does not automatically guarantee that evaluations will be valued. Conversely, respondents believed that legality invites formality and routinization, which can inhibit the effective utilization of evaluation for decision making. Moreover, a few respondents indicated that the concepts, uses, and users of evaluation specified in these directives cultivate a particular evaluation regime with inherent limitations, as discussed below.

Positioning of Evaluation within the Education Decision Domain

This section discusses how evaluation is currently positioned within the Turkish educational decision domain. Drawing on the characteristics of the decision making context and the propositions of Decree No. 652 as discussed above, the current positioning of evaluation in the education domain, including its institutional structures, including existing data systems and human capital, as well as its many definitions, espoused purposes, uses, and objects, are discussed in detail. In doing so, the section argues that the value of evaluation for educational decision making in Turkey is currently challenged by ongoing confusion about and unfamiliarity with the concept.

Institutional structures for evaluation. First, the interviews uncovered the kind of evaluation function envisioned for the education domain by Decree No. 652. Although the Decree mandates internal evaluations that are largely welcomed by the study participants, a majority of government officials shared their hesitations regarding the current role and position of M&E units within the institutional structure of MoNE.

M&E units "at the bottom of hierarchy." Decree No. 652 established monitoring and evaluation (M&E) units under each directorate at MoNE for the first time in the history of the Turkish educational system. Charged to monitor and evaluate curriculum,

instructional materials, processes, and outcomes, almost all government officials

acknowledged the "vital role" these units will play in educational decision making. M&E

officials interviewed for this study reported that they currently monitor the educational

system through centralized data systems and develop periodic reports with data charts

and figures based on predetermined indicators of performance as specified in the

Ministry's strategic plans. While M&E units can initiate their own agenda for evaluation

activities, they are often assigned specific questions and tasks by their directorates.

Despite their efforts to provide systematic information to inform decision making, M&E

officials recognize that it is up to top management as to whether and how this information

will be utilized.

Despite these ongoing activities, confusion among government officials about

how to meet these new evaluation requirements remains. The perception that evaluation

is a new concept and there is a lack of expertise in evaluation greatly adds to this

confusion. Almost every government official indicated that Decree No. 652 created a

systematic framework for evaluation that they "are now trying to fill in," and "since there

is no serious experience in the field, every unit is trying to find its own course of action,"

as one official observed. Despite high-volume, sophisticated, centralized database

systems at MoNE, informants believed that "[t]here is a serious problem in

conceptualizing what we need to do now."

Lack of clarity of roles and responsibilities exacerbates this confusion. The

Decree mandates that M&E units also handle institutional lawsuits, which creates the

biggest obstacle to conducting actual policy evaluations. One official remarked: "A

majority of our work at the Monitoring and Evaluations units now is dealing with

lawsuits, which may be considered as returning to ground zero as far as clumsiness is concerned. This is one of the biggest predicaments we face." Since this predicament will not be resolved any time soon, three officials projected that these units "will get closed" due to systematic malfunction, thus depriving the Ministry of their structural evaluation capacity.

This projection reflects the hierarchical structure at MoNE, as discussed above, where M&E units are now located at the bottom. More than half of the respondents claimed that the institutionalization of the evaluation function is not sufficient to remedy the rampant lack of awareness and internalization of the value of evaluation in educational decision making. Nor does institutionalization guarantee a widespread and effective adoption of the practice. Instead, these respondents imply that mandating evaluation in this context may render evaluation ultimately a mere bureaucratic function. The observation that M&E officials spend most of their time handling lawsuits instead of conducting evaluations led some informants to argue that evaluation's value for decision making is not yet fully understood. This non-association suggests the image of a "multi-headed surveillance system," discussed below.

Silos of evaluation function. Confusion about M&E's role reminded informants of existing mechanisms for oversight in the educational decision domain. Informants mentioned two other major actors and their systems in place to monitor and evaluate educational decisions, leading to more confusion among respondents about what purpose M&E units fulfill.

The first major actor overseeing the implementation and outcomes of educational initiatives identified by respondents is the Ministry of Development. Located at the top of

the hierarchy in setting expectations for the educational agenda as discussed in Chapter 4,

MoD defines each performance target and indicator for education – almost always in

quantitative terms (10th National Development Plan, 2013), "based on and supported by

domestic and international research results," as one official described. The MoD Annual

Programs and Conjectural Evaluation General Directorate is responsible for compiling

monitoring data every three months from the MoNE Department of Strategy to supervise

implementation and budgetary expenditures specified in annual and investment plans.

Although this unit was established within the last few years and acts an "advisory body"

attempting to assess "the big picture," according to one official, it has the power to pull

funding from inactive projects and programs as a reprimand, such that these projects may

never reappear in subsequent investment plans. Despite its recent establishment, a

number of informants felt that MoD's power in monitoring educational activities and

outcomes overrides the role of internal M&E units at MoNE.

The second major actor in guiding evaluation activities in the education domain is

the international donor community, which funds and/or implements educational projects.

Each international project has a steering committee, consisting of government officials

from relevant line ministries (i.e., the Minister, undersecretary, directors, and

bureaucrats), civil society representatives, and international donor representatives, that

meets regularly to review project implementation. The officials interviewed frequently

mentioned the "evaluation function" of such committees in assessing the project's

progress in meeting its goals. In addition to such committees' inherent "evaluative"

oversight, international donors hire external, foreign consultants to carry out project

evaluations. These evaluations are highly esteemed among project officials because they

are found to be "systematic, professional, and independent" and recommendations are believed to be more tangible and useful. The perceived quality of external donor evaluations may threaten the role of M&E units in providing information to project offices.

These external actors provide an "objective eye" for the implementation and goal attainment of educational initiatives, but they make the role of M&E units less clear, since the M&E units share responsibilities and data with multiple other actors. M&E units' internal evaluation service also raised concerns among participants – governmental and nongovernmental alike – about the potential loss of objectivity and credibility of evaluation evidence. A number of informants noted that the Ministry's access to rich educational data might not directly translate into credible, objective evidence in the hands of internal M&E units. According to almost all officials, objectivity should be a higher order aim in any evaluations, yet they are unsure "how can an institution critique its own work honestly?" "Public institutions cannot be critical of their own work by default," a civil society representative argued, and continued:

> The evidence produced by us or others could have better quality than the evidence produced by the Ministry. The reason for this is that the Ministry or other institutions can get really defensive and they can try to create a more positive picture considering their own or institutional image. External organizations can be more judgmental and critical.

As a result, one government official summarized the existing system of educational oversight in the educational domain as a "multi-headed surveillance system," where the M&E unit is believed to be at "the bottom of the hierarchy," according to another official.

Data management systems. While M&E units seek their niche within the educational domain, many informants seem confident about the existence of a sound

database system that could support evaluative studies. "We are one of the most ambitious countries in the world in terms of data collection and management," one official argued, signaling that access to data has never been a problem in the Turkish education system; it is rather proper utilization of that data for a valued purpose that is disputed among respondents.

The discussion about educational data highlighted the role of the Department of Strategy as the central authority for compiling, storing, managing, and reporting educational statistics for domestic (TÜBİTAK, the Turkish Statistical Agency) and international audiences (UNESCO, OECD, and EUROSTAT). Currently, in addition to reliance on international studies such as PISA and TIMMS, officials confirmed that MoNE houses several central data collection systems to monitor the education system (Ministry of National Education, 2009):

- E-school (a national system launched in May 2006 as part of e-government initiative),

- MEBBIS (The Ministry of National Education Information Systems),

- The e-decision support system (funded by the EU under the *Strengthening Statistical Capacity of MoNE Project*),

- The e-performance budget (funded by TÜBİTAK under the *Information Management for Education Finance and Expenditures in Turkey Project*),

- E-statistics (funded by the World Bank under the *Information Management for Provincial and City Directorates for National Education Project*), and

- A performance management system that is still under construction.

Of these, e-school and MEBBIS were most frequently mentioned in the interviews. E-school provides basic education statistics such as enrollment, dropout rate, student achievement, classroom size, and gender, and MEBBIS focuses on the "quality of education," capturing input from school administrators, teachers, students, and parents about their experiences with and perceptions of instruction, learning, and administration. M&E units are instructed to use these statistics compiled by the Department of Strategy to meet their evaluation needs. While some informants were content with these central data systems, a couple of officers noted the inherent limitations of accessing data gathered through a highly structured, bureaucratic process.

In addition to centralized, quantitative data systems, there is a growing interest in conducting fieldwork and collecting raw data from system users – school principals or teachers – through interviews or focus groups. Yet, the officials acknowledged that collecting qualitative data presents challenges, as the participants are not allegedly willing to share their viewpoints unless they are "mandated" by an official letter. One official attributed this to long-standing, centralized decision making, as discussed before, that is seemingly limiting critical outlook among the users of the system.

The predicament of human capital. Evaluation champions – individuals who actively promote evaluation as a decision-making tool – emerged as significant drivers to help M&E units find their niche. A number of informants described how one well-placed individual could put evaluation on the decision agenda as a valuable alternative to the status quo. Several officials believed that top administrators' (i.e., Minister or undersecretary) support for evaluation is key, while others spoke about street-level bureaucrats' power in pushing for an evaluation regime. Regardless of the individual's

position, one government official commented on the significance of personal characteristics and awareness in promoting evaluation:

> We need administrators and specialists who know educational goals very well and can critique educational goals as well. Otherwise, you can give them a couple of data tables; they would just gape and say, 'Well, this looks good.' And your 4-5 million liras are wasted. This is not so uncommon actually.

Drawing on this sentiment, many informants wanted to know whether Decree No. 652 has helped deploy the right talents in pursuit of more effective policymaking. While government officials appeared overconfident in the presence of relevant talent at the Ministry, academics were much more cautionary about existing levels of expertise in evaluation.

Decree No. 652 enabled a new cadre of "young, dynamic, and hardworking" specialists to enter the educational domain, leading almost all government officials to express their hope about the prospects of educational planning and evaluation. These young specialists' "intellectual capacity and thirst" are believed to reduce the influence of political ideologies in educational planning and fill in the perceived gap in analytic thinking among MoNE staff. The background of current M&E staff members includes economics, testing and measurement, English, engineering, math, and public administration. These backgrounds motivated most government officials to attest to the new staff capacity in conducting professional evaluations so much so that evaluation training is not necessary. One official claimed that his staff members are "expert evaluators" because they are "mathematicians, physicists, and engineers." With this cadre of specialists, one official believes, "Everything will work like clockwork."

Contrary to government officials, nongovernmental informants doubted the presence of necessary expertise in evaluation in the decision domain among the new

M&E staff. They specifically pointed to the intricacy of conducting systematic and professional evaluations and expressed concern about the state of staff knowledge and experience. To them, the fact that these new, young specialists can speak English and hold Master's degrees by no means implies that they can conduct quality evaluations. The following comments made by two education professors reveal this sentiment:

> Do we really know what it takes to evaluate a policy? To what extent do we have that capacity? Do we know how the process works? Some people with only teaching background come up to me and say they would like to conduct performance assessments. This is a topic even I am nervous or feel incompetent about. I don't know where we get that courage to think that we can just do it.

> To establish an autonomous, unique and efficient (evaluation) system, we need a cadre of specialists who can collect data, analyze data, and turn these data into decisions in order to sustain the system. Expert staff is of utmost significance here because it is the human capital that manages and improves the system. You know what they say? If you have a good judge, you cannot have bad laws. Our system has changed, but we do not have experts who can fulfill the system requirements.

These informants concluded that the educational domain needs a cadre of specialists trained in educational planning and evaluation who can integrate analytic and technical knowledge and capabilities within Turkey's political, historical, and cultural context to optimize the utilization of domestic and global data. A few of them were skeptical that this is currently the case. What the Ministry has now is a "novice cadre," one academic observed, and continued, "Hence it is not sustainable with the current labor force's capacity and line of vision."

To sum up, the institutional structures for evaluation are not optimally configured to fully build the value of evaluation for educational decision making. While the existing data systems are praised, the new recruits' level of expertise, and sharing responsibilities across departments may undermine the role of M&E units, hence the value of evaluation.

Conceptions and definitions of evaluation. Second, this section explicates the differentiated meanings of evaluation in the Turkish education domain, as expressed in documents and by the respondents (see Table 16). The interviews revealed a common sentiment among participants that there is confusion in the education sector as to what evaluation really means. One official confirmed this confusion and suggested: "Just ask all the directors of M&E Units. They will all give a different answer to the question of what evaluation really is." These evolving conceptions contribute to the limited role and utility of evaluation in the education sector.

Table 16. Different Conceptions of Evaluation in the Turkish Education Domain

Concept of Evaluation	References[1]	Interviews[2]
Quantification	71	25
Performance management	49	19
Inspection and compliance	30	17
Monitoring	24	12
Research and development	23	12
Pilot studies	14	9
Qualitative / Interpretive	13	8
Competitiveness / Materialism	12	6
Internal control and audit	8	6
Critique	7	4
Justice / Ethics	4	4
Procurement	2	2

[1] Total number of times the specific concept of evaluation is explicitly discussed.
[2] The total number of interviews out of 35 in which any concept of evaluation is explicitly discussed.

Quantification. Quantification was the most mentioned and discussed conception of evaluation, signaling the apparent hegemony of counting or empiricism in the education decision domain. One government official's remark summarized the sentiment expressed by many officials: "[Evaluation] used to be called statistics; now it is called evaluation. Nothing has changed though." Several informants frequently and explicitly used such terms as science, objectivity, empiricism, rationality, systematization, statistics,

indicators, evidence-based or data-based, and randomization when they talked about evaluation. A number of informants also equated evaluation with "measurement and testing of student achievement" by mentioning international student tests such as PISA and TIMMS. They believed that quantitative indicators help support Turkey's position in the global context. One official claimed that, "We need this [kind of] data to demonstrate Turkey's compatibility with the global world, and we adhere to the rules of modernity." They attach the value of evaluation to its production of statistics and support of "rationality," upon which "sound decisions" can be made.

As in the interviews, evaluation is overwhelmingly portrayed as an approach characterized by quantification and rationalization in official policy documents. MoNE's first strategic plan defines evaluation as follows: "The measurement of results compared to predetermined goals and objectives, and the analysis of these goals' and objectives' relevance and consistency" (Ministry of National Education, 2009, p. 207). This plan and MoNE's annual activity reports demarcate every goal and performance indicator in quantitative terms (see Ministry of National Education, 2011, 2012, 2013). Overreliance on quantitative data was also observed in the final report of the EU's Support for Basic Education Programme (SBEP) (2002-2007): "The dominant research tradition is highly quantitative but now needs to be much more qualitative oriented if it is to take a much needed and important role in informing major national and local policy and strategic decision" (p. 42). The existence of quantitative data systems introduced before may also contribute to this rampant association.

Similarly, every national development plan specifies educational targets in quantitative terms. Each education indicator is operationalized in greater detail and

assigned a budgetary figure, paving the way for systematic monitoring of public expenditures. For example, the target of achieving 100% access to primary education is operationalized down to the number of classrooms needed and expenditures required to buy classroom materials (Ministry of Development, 2006, 2013). These documents may contribute to strengthening the connection between quantitative measurement and evaluation in the education domain.

While many informants equated evaluation with quantification, a few academics and government officials expressed displeasure about this. According to these informants, overreliance on statistics and counting may sabotage efforts to produce useful information for decision making because quantification breeds a culture of monitoring in which the question of "so what?" is often overlooked. One academic specifically argued that quantification reinforces the "linear, reductionist, and false-promising" logic of the production function in education. A number of academics believed that the logic of quantification is borrowed from Western education systems that promote the utilization of quantitative indicators through global student achievement tests such as PISA and TIMMS and critiqued the quantification of every educational outcome and output. These informants feared that overreliance on numbers may develop into an evaluative culture in which even qualitative findings are quantified to the extent that creativity and critical thinking are compromised. A few government officials acknowledged that recent national education reports did not provide any explanation beyond mere tables and charts (see Ministry of National Education, 2011, 2012, 2013). Similarly, one government official thought that the M&E units acted as if they were statistics departments, thus the information provided is not always useful. Another official remarked:

> [Evaluation] is perceived as a luxurious term in Turkey. It is professional work. If you cannot sustain its professionalism, then you start dealing with statistics only. The most dangerous thing is when people get used to something. If it is a good custom, it is good. But if it is bad and it has become your culture, then you have a problem.

Calling the trend of quantification in Turkey "scientific control," such informants not only illustrated evaluation's forced-fit with quantitative rationality but also highlighted the more nuanced nature of evaluation.

Performance management. The terms "performance management" and "total quality assurance" were frequently and interchangeably used when speaking of evaluation. According to a number of informants, performance is monitored to ensure "quality in education" so that significant differences between schools can be reduced to provide equitable education for all.

The review of official documents suggests that this association entered the Turkish education domain with the passage of PFMC Act 5018, as discussed before; it requires the use of performance standards to improve efficiency in public service delivery. To this end, organizational structures of many line ministries have been duly reformulated to create effective policies based on performance indicators. The accompanying mandate for strategic planning routinized the link between performance management for fiscal control and evaluation for budgetary allocation (see Ministry of National Education, 2009).

Despite its widespread use by government officials, a majority of the nongovernmental respondents expressed hesitation about the use of performance management in Turkey due to lifelong tenure in governmental positions. Referring to the logic of performance management in the United States, two academics specifically

wondered how the Turkish educational system could ever mandate the use of

performance standards with teachers whose job is secured for a lifetime:

> We do not have contract-based employment. Everyone is a civil servant. As a result, you cannot fire someone from her/his teacher position or even civil service altogether. The courts would never give you the permission. Even if this person is deadly sick, psychologically disabled, s/he still carries the title of a teacher. Am I being clear? There is a significant disconnect between our dominant, established mentality and implementation culture and the concepts and processes we have borrowed or [are] aspiring to borrow.

> People's salaries vary in accordance with their performance in some foreign countries. Plus, teachers do not get appointed. 'The state has got my back. I am already appointed. Oh, I am sitting pretty!' is a common dominator in our system.

A couple of other informants reiterated that this growing culture of performance

management greatly contributes to overreliance on quantification and rationality in

service delivery and goal attainment.

 Inspection (denetim & teftiş) and compliance. A number of officials and

academics associated evaluation with inspection and compliance, based on the annual

school and teacher inspections conducted by the School Development and Quality

Bureau under the Board of Inspection. According to these informants, evaluation is a

process of verifying that educational activities are being implemented in accordance with

the strategic and/or national development plans. Noncompliance may result in sanctions.

For example, schools that do not comply with standards might be closed; inactive

projects might be discontinued.

 However, other informants questioned the association between evaluation and

inspection, arguing that inspection has negative connotations, which are also then

associated with evaluation. One academic cautioned that, "The perception that evaluation

is not a tool for improvement, but for finding out failures has now been instilled in

people." One government official confirmed that school principals or teachers often think

they will lose their jobs due to "evaluation" results based on inspections, so they often do not welcome the idea of evaluation.

Monitoring (izleme). Overreliance on quantitative performance indicators led many informants to associate evaluation with monitoring. The word "monitoring" (*izleme* #643) was literally used more frequently than "evaluation" (*değerlendirme* #491) in the qualitative interviews. Routine monitoring data available through the MEBBIS and e-school systems were often discussed in connection with evaluative activities, although officials acknowledged that monitoring an educational system as large as Turkish one is not an easy task.

Research and development. Evaluation and utilization of evidence reminded some informants of research and development (R&D). A number of informants considered academic studies, more specifically master's theses and dissertations, as "evaluative" documents that could provide sound information to design and improve policy alternatives and decisions. According to these informants, research and evaluation employ the same methods and serve the same purpose. Hence, one government official said, "Evaluation, to me, is R&D in a way," and one civil society representative could not see any difference between research and evaluation in practice.

Pilot studies. The qualitative interviews uncovered a unique category of evaluation-related meaning. Evaluation was often equated with pilot studies – the ex-ante studies – conducted before large-scale policy implementation. In other words, evaluations are mostly disguised as pilot studies. These pilot studies were presented as the best approach to incorporating evaluative information into decision making, legitimating policies' existence and continuation. Government officials explained that pilot studies are

often conducted in randomly sampled schools or classrooms across the country, using mostly quasi-experimental designs; they provide information on implementation and potential policy outcomes that can justify system-wide adoption and replication. One official described the utility of pilot studies by referring to the School Milk Project – mentioned in Chapter 1 – because of which some students were hospitalized. If a pilot study had been conducted, this official claims the incident would not have happened. Officials believe that pilot studies not only test various ideas and alternative policy formulations, but also identify emerging needs and contextual challenges in implementation.

Despite the widespread use of pilot studies, a number of informants critiqued their use, arguing that the educational authorities do not share the results of these studies before scaling up the intervention or program. Another critique concerns the pilot studies' gradual replacement of ex-post evaluation activities. A couple of informants think that pilot study results are considered definitive, such that policies are not subject to ongoing systematic evaluations.

Interpretative and qualitative. The discussion of evaluation enticed some respondents to discuss the implications of particular methodologies. A few informants mentioned qualitative methods – mostly interviews and focus groups – employed in some studies. By virtue of their inclusiveness, qualitative methods were highly esteemed by these informants. Beyond immediate methodological concerns, a couple of academics questioned the epistemological assumptions of particular evaluation methodologies. For example, one academic challenged the overreliance on quantitative indicators as

discussed above and argued that a loss of interpretivism may be detrimental to information produced to inform decision making.

Competitiveness and materialism. Six informants associated evaluation with the competitive market economy and materialistic culture of the Western world. To this line of thinking, evaluation fosters competition among schools, teachers, and students who race to outperform their peers according to preordained criteria and standards. It also helps the entire country assess its position in the global knowledge economy. This leads to the development of a materialistic culture in which outcomes are tangibly articulated, obtained, increased, and manipulated, based on which schools or students "advertise" their achievements in test scores. By seemingly facilitating a subtle smooth transition to a free market economy, these informants indicated that evaluation is believed to threaten local practices, which tend to harmonize material and non-material processes.

Other. In addition to concepts mentioned above, a number of other meanings associated with evaluation emerged from the data. Some informants spoke specifically about the auditing function of evaluation in ensuring efficient use of public resources – currently, the responsibility of the Office of Internal Control and Audit. Others associated evaluation with criticism, judgment, and critical thinking, all of which ascribe negative connotations to evaluation. Four informants described evaluation's potential "justice restoration" function by holding government officials "conscientiously accountable." Last but not least, evaluation reminded a couple of informants of the procurement of educational services and products.

In short, the many meanings and differentiated conceptions of evaluation tend to confirm the lack of history with the practice of evaluation in the Turkish education decision domain, and explain why the M&E units are endeavoring to find their niche.

Purposes of evaluation. Despite the conceptual plurality as discussed above, there was greater consensus among informants in terms of evaluation's main purpose (see Table 17). In almost all interviews, evaluation was treated as a retrospective tool to assess goal attainment at the end of a policy cycle. A substantial number of informants believed the purpose of evaluation was to determine the extent to which the educational initiative – policy, project or product – produced the intended outcome or simply to answer the question "Did it work?" Goals, results, impact, achievement, outputs, and inputs were terms frequently used in connection with this common purpose of evaluation. Conversely, very few informants spoke specifically about understanding the process of policy implementation in their definitions of evaluation or critiqued the heightened emphasis on results.

Table 17. Purposes of Evaluation for Turkey's Educational Decision Making

Purpose	References	Interviews
Goal attainment	35	21
Implementation/process	11	8

Below are the description of specific purposes under goal attainment and implementation highlighted by participants, some of which were based on actual reports of evaluation and some that were as of yet unrealized, but anticipated. Given the evaluation's association with quantification and performance management, evaluations' anticipated role of assessing goal attainment is justifiable.

Accountability. Evaluation was often closely associated with the purposes of accountability and transparency. But curiously, several informants believed that evaluation was the effect of accountability, not the other way around: when the principle of accountability is the governing paradigm in governmental affairs, and when individual and institutional conduct is structured to sustain that principle, then evaluation's value in demonstrating that "The [European] King is naked" is recognized and justified. One academic described this connection as follows:

> Politics needs to be transparent. One of the features of a social welfare state is to hold people accountable by creating a transparent system where no decision is made behind closed doors. Everything has to be transparent. Then monitoring and evaluation can be associated with accountability and transparency, which appears to be highly unlikely for Turkey at this moment.

But even though accountability and evaluation are respected in Turkey, these informants acknowledged that the system could not hold people accountable based on evaluation findings because of the norm of lifelong incumbency that inhibits intentional behavior change based on evaluation. One civil society representative characterized the current mentality as: "It does not matter whether we get happy or regretful [at the end] we will still do it anyway." In the absence of structures rewarding change as a result of accountability measures, the value of evaluation is diminished.

Concerns about accountability exposed the interplay between the fear of criticism, as discussed above, and the right to information. A majority of nongovernmental informants believed that government officials' fear of criticism leads them to share only limited information regarding the results of programs and policies. Even though a law on the right to obtain information was passed in 2003, no major change is believed to have occurred in the educational domain that remains as black box to many. In return,

nongovernmental informants believe that this further diminishes the transparency of governmental endeavors.

Educational authorities, on the other hand, indicated that they are trying to protect the institution's image, which is easily deteriorating to the extent that officials are often "worn down." They recognize the limitations of sharing information related to their internal work, which impacts the rest of the Turkish population. Yet, they think that if they share all the information they have, it may hurt more than it benefits, hence "they (figuratively) hide some information." One government official summarized this common sentiment among officials: "Monitoring and evaluation units conduct studies that require technical expertise; they are not charged with marketing and advertisement. It is not their job to broadcast the results." One official said: "Would it be better if we shared the information? [Yes,] I agree, but people whom we shared the information with should make constructive criticism; they should not be destructive. Otherwise, we cannot have productive discussion." This sentiment is tied back to perceived fear of criticism among officials in that their lack of trust in how non-officials will use the data is self-reinforcing.

Nevertheless, some informants believed that the recent public administration reform will gradually alter the prospects of sharing information. PFMC No. 5018's mandate for strategic planning and performance-based budgeting for public institutions was believed to infuse the principle of accountability into public service delivery. In addition to improvements in the public sphere, civil society organizations were reported to be taking more responsibility in filling the gap in information regarding government initiatives. These developments were expected to encourage educational officials to be more open and transparent, further highlighting the value of evaluation.

Learning and improvement. While government officials interviewed for this study did not mention accountability as a potential purpose of evaluation, they focused more on the learning aspect of evaluation. To them, the purpose of evaluation is problem-solving and sense-making, uncovering the weaknesses and strengths of the system so that necessary programmatic and policy modifications and improvements can be made. Their definitions of evaluation strictly focused on "revising and modifying policies and programs" based on evaluation results. "This is the only way [evaluation] has to go," one official stated, and continued, "After all, you cannot keep sweeping all negatives under the carpet. It would eventually get full." Yet, they did not provide any specific examples of how current policies and programs have been modified or improved based on evaluation feedback. For example, many informants, including government officials, wondered if the recent and controversial law known as 4+4+4 is currently being evaluated. A couple of informants mentioned how the Ministry changed the age requirement for starting school from 60-month-olds to 66-month-olds under this regulation. Although a few educational officials reported having consulted with the civil society organizations to discuss alternatives and make necessary and timely changes to this policy, one official claimed that the change probably was due to overwhelming media coverage and critique, not necessarily to credible evaluation information, which is believed to be currently nonexistent.

Policy development. A number of informants mentioned the developmental purpose of evaluation. Since the qualitative interviews confirmed that formal policy and program evaluations were rare in the educational decision domain, they implied that the closest educational officials have been to conducting and utilizing evaluations has been

policy development. Pilot studies and basic research, as discussed before, as well as needs assessment and situational analyses were considered *a priori* evaluative activities that provide evidence for developing policies or addressing emerging needs.

In sum, evaluation was commonly defined as a tool to assess policy or program outcomes with a purpose to restore or strengthen accountability. The learning aspect of evaluation was also mentioned as a way to improve policies by learning from mistakes. Since the notion that evaluation is a relatively new term is commonplace among the study participants, most of the discussion around what purpose evaluation serves was theoretical at best. This observation explains the M&E units' pursuit of niche within the education domain in that actual accounts of evaluation studies were rarely shared.

Uses of evaluation. Despite the substantial consensus on evaluation's emergent evolution in the Turkish educational domain, a majority of informants wanted to use evaluations instrumentally for decision making, drawing a direct, correlational line between results and action. They placed high significance on the need for evidence to alter the educational domain's resemblance to a "jigsaw puzzle" of free flowing educational initiatives, offering evidence as an antidote to political interests. In particular, a majority of government officials affirmed their belief in evidence-based decision making, stating that continuous improvement is a prerequisite for a legitimate government. Another official indicated that the Ministry's principle is "right data, right decision." One common perception among government officials was to "keep up with the developed countries" that also use evidence for their decision making.

But a few informants expressed hesitation regarding the instrumental use of evidence for decision making. One government official did not expect that evidence

would be used 100% of the time in the education domain, hence he suggested offering

multiple policy scenarios and alternatives based on available data to expand decision

makers' horizons. An academic specifically pointed to a lack of participatory decision

making that yields incomplete evidence, hence inaccurate reality. The academic also

revisited overreliance on quantitative indicators and argued that:

> I have some reservations about data-based decision making… People tend to
> work with quantitative indicators mostly; they try to develop policies based on
> test results for example, and we know that test results do not tell everything about
> students... a culture of teaching-for-the-test or study-for-the-test develops as a
> result. I think data-based decision making is important but we need to discuss
> what the data are supposed to look like and how diverse they need to be.

Similarly, civil society representatives leaned more towards the conceptual use of

evaluations in their effort to eventually increase the role of evidence in decision making.

One representative indicated:

> MoNE is not a monolithic institution. It is still an impact if you can influence
> decision makers' behavior or attitude, change the way they think about or
> approach policymaking just even a little bit. Thus, we need to define impact in a
> much larger framework. Otherwise, it will be unfair to civil society organizations
> or organizations that deal with public policies, education policies because we need
> to underline the fact that they are not decision makers. Nor do they have any
> power of mandate.

In sum, a subtle tension emerged regarding the instrumental use of evaluation

findings directly for policy or program decisions. Given the hierarchical and politicized

nature of educational policymaking in Turkey, nongovernmental participants appeared to

favor the conceptual use of evaluation findings to gradually change the mindset of

decision makers over time. Yet, some officials seemed hopeful and expectant about the

instrumental use so that data-driven "right decisions" can be made. This tension further

reflects the confusion about what evaluation can really offer in the Turkish context.

Objects of evaluation. The discussion of purposes and uses of evaluation for educational decision making highlighted some sample evaluation studies. Throughout the qualitative interviews, several objects of evaluation were discussed, among which projects were most frequently quoted and programs were least mentioned. Table 18 summarizes the objects of evaluation mentioned in the interviews and documents reviewed. The table illustrates and further confirms the involvement of donors in the Turkish education sector. Evaluation was frequently discussed in the context of donor-funded education projects, which introduced the concept and practice into the sector as discussed before.

Table 18. Examples of Evaluands Discussed During the Interviews

Evaluand	Examples
Project	Monitoring of Vocational and Technical Schools Project (MTEM); Quality of Vocational Schools Project; Strengthening of Early Childhood Education Project; Fatih Project; Conditional Cash Transfer; School Milk Project; Violence Against Children Project; Democratic Citizenship Project; Strengthening Human Capital Project; Supporting Basic Education Project; METKEP; METEG; Drop-Out Rates Project
Product	Curriculum, lesson plans, textbooks, student portfolios, educational conferences and workshops
Policy	4+4+4
System	PISA and TIMMS scores, MEBBIS and e-school data
Procured Items	Instructional technology; maintenance of school buildings and school supplies

The overwhelming referrals to projects as evaluation objects during interviews signaled the extent to which evaluation is incorporated into decision making, hence its assigned value. One official indicated that since the introduction of project management around the 1990s, "countless projects" – mostly internationally funded and delivered – were implemented in the education domain, some of which have been introduced before.

Another official shared, "There is nothing that has not been tried, absolutely nothing" in the context of such projects.

On the one hand, the nature of projects was seen to enable the conduct and use of evaluation. Project officials expressed that since projects are short-term and targeted with a specific budget, a discussion of their impact becomes inevitable. Common providers of project evaluations were noted to be the bilateral and multilateral funding organizations (World Bank, UNICEF, the EU, UNDP, etc.). During the interviews, these organizations were frequently mentioned as powerful actors, and by virtue of their perceived independence and professionalism, their evaluation services were highly esteemed, thought to produce tangible and practical feedback through a systematic process. One official argued that, "The EU projects are spotless in this sense. They are designed to impact policies with their big budgets."

On the other hand, a majority of nongovernmental informants expressed confusion about these projects' real impact in Turkey, hence the image of "project dumpster" introduced above. These informants had the impression that these international projects did not serve any valued purpose other than providing financial incentives. A few academics who served as external consultants for some of these projects indicated that they had not been implemented in a "healthy" manner, but did not elaborate.

Only one government official out of 35 interviewees talked specifically and intentionally about the logic of programming and how the educational domain in Turkey should focus on programs rather than short-term projects. This official defined programs as multi-faceted, longer-term initiatives that are more inclusive and comprehensive than projects. The official claimed that programming logic would partially remedy the

education domain's image of "project dumpster" and assist incorporating an evaluative mindset into decision making.

In sum, education projects, mostly funded by Turkey's development partners, overwhelmed the conversations about sample evaluation studies. A lack of discussion on nationally-driven and designed policies and programs reflects the influence of donors on the emerging practice of evaluation in Turkey.

Future of Evaluation: "25% and loading…"

Given the lack of history with the concept and practice of evaluation in Turkey as noted by the participants in this study, the future of evaluation as a form of inquiry was contested. Despite the institutionalization of the evaluation function at MoNE, confusion about what evaluation really means and what it does really offer to decision making confirmed this finding. This section outlines the opportunities and barriers to building and sustaining the value of evaluation for educational decision making in Turkey.

Too radical to be true. Overall, Turkey's bureaucratic culture and the unmanageability of the growing education system is supposedly the most debated issue in the Turkish educational decision domain since the 1980s, so much so that Dinçer's first act was to put a dramatic stop to these problems by initiating the design and passage of Decree No. 652. The tremendous growth of the educational system over the years with more schools, teachers, and students necessitated a new, more effective, responsive, and manageable approach to administration (Urgent Action Plan, 2003). This restructuring of the central education administration has been mentioned in government and national development plans several times (Government Action Plan, 2002, 2011; Ministry of Development, 2006, 2013). Specifically, Urgent Action Plan (2003) prioritized the

restructuring of the education system, which at the time comprised 50 different units with over 5,500 staff members – a "clumsy bureaucracy" that made it difficult to provide effective services. Hence, educational governance reform was long considered "essential," "inevitable," and "imperative."

Nevertheless, a number of informants indicated that the change in educational administration introduced by Decree No. 652 was too radical and too fast in execution, which left little room for thorough deliberation and reflection, thus posing challenges to the future valuation of evaluation. The size of the education system led an official to argue as follows:

> The educational decision domain resembles a multi-trailer truck. The central administration is on the steering wheel with the provincial units located at trailers. It is both difficult and hazardous to control, maneuver, and operate these multiple trailers all at once. When the central administration steers or rotates sharply, rollover is highly likely.

Thus, change making should be "evolutionary, not revolutionary" in this context, one academic suggested, in order to ease the process of adaptation and increase the chances of internalization. Since the execution of reform movements is believed to be radical in Turkey, the public has developed a distaste or even apathy toward reform movements, as two academics warned. One government official stated, "You need to manage the perception [of people]. We are talking about 17 million students. Change making and innovation will not be easy. You cannot just say 'But I did a needs assessment'; policymaking or programming does not work that way." As a result, evaluation's conceptual and practical anchoring in the education decision domain is believed to change depending on political fluctuations in the decision domain.

The predicament of organizational memory. Dinçer's radical change is believed to have affected the institutional memory of MoNE, and organizational memory is found to be paramount in affecting the future value of evaluation. One government official remarked:

> It is so hard to erase the things that you already know. It is hard to get rid of evidence. We need to be a 'forgetting society' before we can be a 'learning society.' One can easily be trapped with the already learned schemes. Then, you tend to think, 'Nothing good will become of him.' For example, you have a staff member who constantly refers to the historic facts and who just does not wish to change his mental schemes. So then it does not matter really that you changed the system unless the individuals have changed.

Two camps emerged during the interviews regarding the impact of Decree No. 652. On the one hand, the new M&E system brought in new specialists who are believed to lack a working memory of how things work at the Ministry. One academic remarked:

> There is a missing piece here. There are staff members that had worked for the Ministry for years. Their experience is really important. All of a sudden, we went through this spectacular change. Specialists with master's or doctoral degrees are brought in. But they do not know how the system works. We needed a bridge [so that] they can combine theory and practice and strive to do better. This change should have been incremental. They removed the directors in each unit all of a sudden and brought new ones in. The bridge has broken. Something now is missing.

On the other hand, some think that the young specialists will help create a solid, sustainable institutional memory along with a new system. This new institutional memory would help solidify and strengthen the place of evaluation in the system. Respondents affirmed that institutional memory has a subtle yet very powerful influence on the kind of evaluation concept and system to be internalized. That is why some believe that it is an extremely delicate matter to develop a particular mindset early on in the process with these new specialists, whose minds are fresh and do not carry the burden of past practices.

Infant yet promising. Even though Decree No. 652, which is commonly seen as radical, officially put the evaluation function on the education decision agenda, a majority of informants believed that a complete adoption of evaluation as a decision-making tool in the Turkish education domain will be a work in progress. One official called the current progress as: "We are at 25% and loading…" This is largely due to the reported confusion about the practice of evaluation that is still debated. One government official said:

> M&E units do not have a history in the educational domain like other directorates…M&E is a new concept in Turkey, in the Ministry of National Education…We kicked off with the Law No. 5018. Internalization of the term will be different across departments, however.

Given the difficulty in changing people's perceptions of evaluation, officials are cautious about their projections. Three officials commented on the progress of evaluation becoming institutionalized at MoNE:

> It is a transition period. We can say that we are crawling now. Hence, we cannot say that [evaluation] evidence is provided for decision making yet and this is normal and expected. What else is better than this? As I said, when all the pieces of the puzzle will fall into place, the system will improve itself.

> Every change movement is a bit painful whether we like it or not. Some people will oppose the change but it looks like we have partially broken this clumsy bureaucratic system with restructuring. Yet, full establishment of new systems takes time. It cannot be accomplished, say, in three days.

> The resistance of those who are familiar with or benefit from status quo and those who simply fear or cannot estimate the results of a new system although they are largely discontent with the current system will be the biggest obstacles against the change the Decree 652 brought about. This is indeed the destiny of those who dare make radical changes. Change making is not the job of those who fear resistance.

Despite evaluation's infancy, a majority of the informants were hopeful about the future of evaluation in the Turkish education decision arena. One academic with a background in government shared:

The idea that we are incorrigible is in our culture. But this mental schema will disappear with globalization. The guy who invented the total quality assurance also went through a long process of trial and error. This is a question of incremental change.

One government official hoped: "We named our baby (M&E), now we need to wait and see how it will grow up."

Conclusion

To conclude, a conceptual framework is presented as a visual guide to understanding the constructs and tentative relations among them that emerged from the findings (see Figure 15). This framework is not meant to represent a theoretical relationship among suggested phenomena, nor does it intend to prescribe any specific course of action to strengthen the value of evaluation for educational decision making in Turkey or elsewhere. It aims to contribute to the discourse about the role and utility of evaluation in public policymaking through empirical inquiry. The framework is depicted as a tree – a popular metaphor to represent complex phenomena in the field of evaluation, as in Alkin's (2004) *Evaluation Roots* – to capture the interconnectedness among principal constructs of interest and highlight antecedents to and precedents of their complexity.

At the bottom of the tree, the framework draws attention to two antecedents that deeply influence the value of evaluation: *underlying principles* and *decision makers* who make meanings and uses of these principles. Turkey's rapid economic development coupled with her desire for modernization suggests particular educational governance that is grounded on the principles of good governance, results-based management, efficiency, effectiveness, accountability, and transparency. While Turkey's economy is gradually pushing to highlight the value of evaluation as a decision-making tool, the extent of

compatibility, contestations, and negotiations between the so-called Western culture and Turkish ways of thinking and living will breed new possibilities and generate new tensions or opportunities for the value premises of evaluation for educational decision making. At this juncture, the role of national decision makers and international actors will be influential in making meaning and use of those opportunities and tensions.
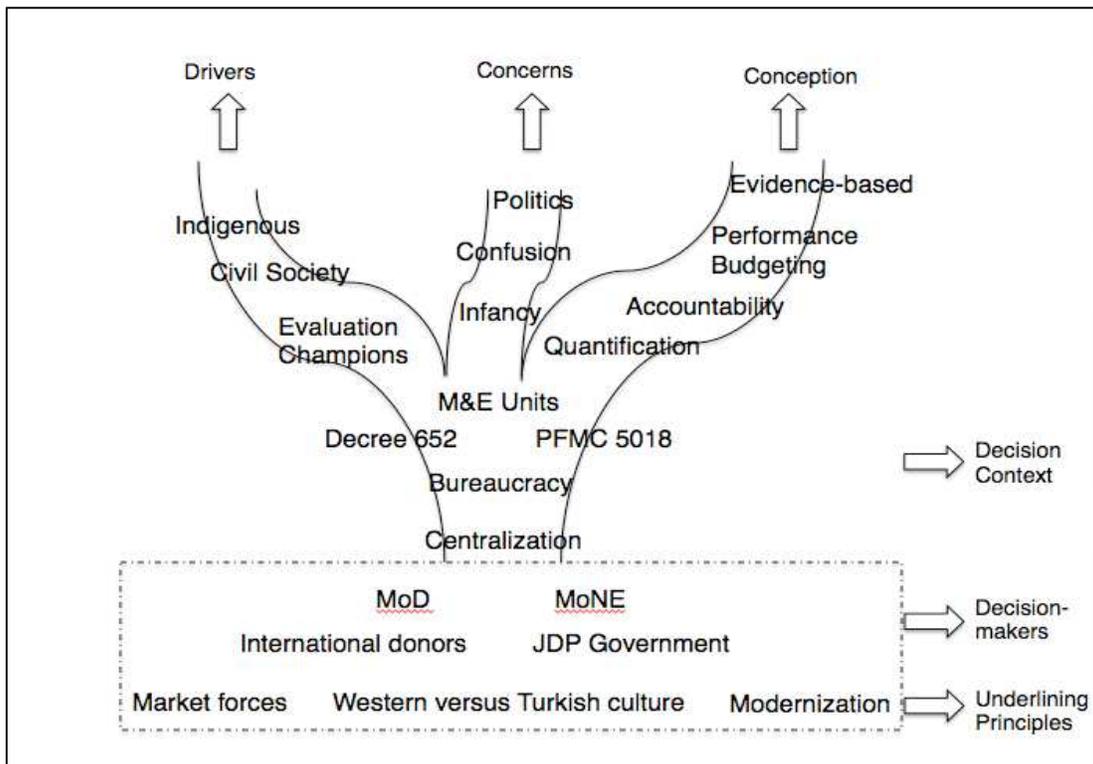
Figure 15. A Conceptual Framework Depicting the Value of Evaluation for Turkey's Educational Decision Making

The body of the tree captures the principal constructs of the educational decision-making context – centralization, bureaucracy, PFMC No. 5018, Decree No. 652, and M&E units – upon which the value of evaluation is currently grounded, envisioned, and established. Highly centralized decision making and heavy bureaucracy constitute building blocks of the education domain through which anchoring of the evaluation function and mentality is mediated. Two significant government acts—i.e., PFMC No.

5018 and Decree No. 652—represent organizational support structures that eventually enabled the establishment of Monitoring and Evaluation units at the Ministry of National Education. These units' mandate for internal evaluations and coordination of external evaluations served to shape the nature and extent to which evaluations are proximate to educational policies and programs.

The branches of the tree reflect the precedents of the value of evaluation: *concept*, *drivers*, and *concerns*. Represented within the far-right branch of the tree are constructs – quantification, accountability, performance-based budgeting, and evidence-based practice – related to current, major conceptions of evaluation in the educational decision domain. Findings suggest that quantification is now deeply rooted in the Turkish imagination as the main connotation of evaluation, closely associated with a desire to strengthen the principle of accountability in educational planning and programming. The momentum for performance-based budgeting is the consequence of quantification and accountability, further elevating evidence-based policymaking as a common denominator for conducting evaluations in the decision making context.

Represented on the far left branch of the tree are drivers that will continue to highlight the role and utility of evaluation, namely evaluation champions, civil society, and indigenous practice. Evaluation champions – either young, educated bureaucrats or top managers such as the former Minister of Education – are likely to bear the responsibility for promoting evaluation as a viable tool and form of inquiry on the decision agenda. Civil society indicates the form of evaluation entrepreneurship that is relatively devoid of political agenda, supports citizen-focused service delivery, and advocates for participatory planning. Most importantly, these entrepreneurs are called on

to help contextualize, localize, and indigenize the concept and practice of evaluation that is reportedly needed in the Turkish context.

The middle branch of the tree reflects concerns – infancy of the system, confusion about evaluation, and politics – associated with the sustained value of evaluation for decision making, tempering the leading conceptions of evaluation and drivers for internalizing evaluation's offerings for educational decisions, hence its location in the middle of the tree. A lack of history with the concept of evaluation results in an infant system that is still emerging despite all the structural adjustments and reconfigurations. This emergence is affected by widespread confusion about what evaluation really means and how it should be conducted and utilized for educational decisions. Findings suggest that political forces such as change in educational administration, trends in globalization, or societal dynamics will determine if and how evaluation will flourish. Since evaluation has been brought in relatively radically, the public is suspicious about its likely use and potential for success, as with most of the reform agendas in Turkey that have been introduced overnight. Reformist agendas in Turkey are doomed to change significantly over time, perhaps losing novelty and essence from their original version. The future of evaluation and what it could offer to educational decisions will be determined under what umbrella reform agenda it has been brought in under.

The constructs of indigenous, politics, and evidence-based practice are curiously located at the top of each branch. As the tree of evaluation's value grows, these concepts will be entangled and may give rise to serendipitous roles to evaluation in Turkey. Especially the interaction between the Western construct of evidence-based decision

making and indigenous, local ways of governance will be mediated by the political

discourse of the time in Turkey.

Chapter VI: Conclusion

This study is the first empirical account of the value of evaluation for Turkish public policymaking, specifically in the educational decision context. The bourgeoning literature in evaluation in the global North has overlooked Turkey in its case selections, and Turkey's academic radar has neglected the value of evaluation as a decision-making tool for educational policies and programs. This is a lost opportunity, this study argues, as Turkey may reveal more about the implications of the cross-cultural transfer of evaluation systems and practice. To recover this opportunity, the study answered four inter-related research questions:

1. To what extent and in what ways is evaluation utilized in the Turkish education decision domain?

2. How is evaluation conceptualized in the Turkish education decision domain?

3. What is the need and capacity for evaluation in the Turkish education decision domain?

4. What contextual factors facilitate or hinder the value of evaluation for Turkey's educational decision making?

This chapter begins with an integration of the study's major findings to provide the foundation for a later discussion on research and practice in the field of evaluation. To do so, the chapter first synthesizes the findings by each sub-research question in Table 19. The immediate discussion following this table will explicate how key findings from this research relate to arguments in seminal literature on the value of evaluation. The chapter will conclude with recommendations for future research and practice.

Table 19. Integrated Summary of Research Findings for Each Research Question

Research Question	Evidence from the case of Turkey
1. To what extent and in what ways is evaluation utilized in the Turkish education domain?	Despite a legal mandate for evaluation practice thanks to Decree No. 652, as well as integrated, full-fletched database management systems and educated staff members, evaluation practice does not reflect a stable character in the education sector. The role and utility of evaluation is mostly materialized and recognized through ex-ante pilot studies based on which the system-wide adoption of educational initiatives is legitimated. International donors continue to fund external evaluations of their educational projects, and domestic think-tank organizations endeavor to inform public of the outcomes of policies based on external research.
2. How is evaluation conceptualized in the Turkish education domain?	Evaluation is still considered an emerging construct in the education sector, and a lack of history with the concept and reported confusion about what the practice really entails are believed to contribute to the frequent association of evaluation with quantification. More importantly, the analysis of findings exposed an emerging tension regarding the meaning of evaluation. While reliance on quantifying educational processes and products is welcomed by some to clearly demonstrate Turkey's alignment with the developed world, some problematized this association by identifying the neoliberal underpinnings of practice and called for a more culturally nuanced construction.
3. What is the capacity and need to conduct and utilize evaluations in the Turkish education domain?	The findings suggested a subtle hierarchy between the two significant elements of evaluation capacity for the Turkish case: legal environment and human capital. First, the study notes reservations about the legal enabling environment for evaluation, which is commonly noted in the literature among the most important factors to establish the value of evaluation for decision making in the developing world. Second, the study spotlights the important role played by evaluation champions in advancing an evaluation agenda in a context where it is historically immature and puts forward the argument that it is the human capital that *enables* the environment (either legal or political) for evaluation to flourish.
4. What contextual factors facilitate or hinder the value of evaluation for Turkey's educational decision making?	The findings pointed to two contextual elements that define the current borders of the evaluation concept and practice in Turkey: the politics of educational policymaking, and Turkey's rapid economic development. The study exposed a tension between these two elements. While politicization of educational policymaking is putting evaluation under dimmed lighting, rapid economic development serves to highlight the value of evaluation as a decision-making tool.

Discussion and Implications of Findings

This study set out with the expectation that, given her famous geo-political position, and socio-economic and historic connections with the global North and South, Turkey would present an opportunity to test assumptions in the field of evaluation. In line with this expectation, the results revealed that, although Turkey's surge of development and modernization serves to highlight the value of evaluation as a decision-making tool, evaluation remains in limbo in the educational decision domain due to resistance generated primarily in the political and cultural context. The study highlighted the role of evaluation champions as the most significant factor to pull evaluation out of obscurity and help localize the concept and practice for Turkey, which is believed to be a work in progress. To develop this argument, and in light of the summary table above, this section discusses the major findings that surfaced during interviews and document review for each sub-research question and puts forward how these findings relate to literature in the following order: context for evaluation, concept of evaluation, need and capacity for evaluation, and current utilization of evaluation.

Unpacking the context for evaluation in Turkey and beyond. In line with the central premise of this study, that is, that context matters in articulating, negotiating, bordering, practicing, and utilizing evaluation as a form of inquiry for decision making, first understanding how Turkey's national context interacts with evaluation – which factors inhibit or facilitate the value of evaluation for educational decisions – is necessary. This study did not seek to provide a complete account of policymaking mechanisms in Turkey, but rather to present a set of approaches that permit a promising understanding of how evaluation's value is mediated by the current context with relevant

policy actors, their needs, interests, and overall decision structures. The findings pointed to two contextual elements that define the current borders of the evaluation concept and practice in Turkey: politics of educational policymaking, and Turkey's rapid economic development. The study exposed a tension between these two elements. While politicization of educational policymaking is putting evaluation under dimmed lighting, rapid economic development serves to highlight the value of evaluation as a decision-making tool.

Political context. The findings indicate, first and foremost, that the role of evaluation – including its definition, purposes, questions, design, methods, uses, and influence – in the Turkish education system must be situated within the role of education in Turkish social and political life. In other words, the findings suggest that the foundations of the education system in Turkey create an opportunity structure within which the foundations of evaluation in the educational decision domain are negotiated, established, and modified. As demonstrated by the depiction of the education domain by the informants and policy documents, education has been conceived and utilized in Turkey as an instrument to build national identity, install accompanying state ideology, and transform the entire society in a particular way ever since the establishment of the Republic (Şişman & Turan, 2003). Education's ascribed role for social transformation necessitated hierarchical decision making centered under the administration of the Ministry of National Education (MoNE) so that the power of control over education and dissemination of state ideology as a single paradigm is consolidated and uncompromised (Turan, 2000). Despite Decree No. 652's radical attempt to decentralize the system and facilitate inclusiveness in decision making (Aydagül, 2013), the philosophical and

strategic incoherence of Turkish educational policies and programs due to fluctuating

agendas and dominant ideologies has been troubling, hence widely attracting criticism.

The following remarks made by Turkish scholars illustrate the common foundations of

the educational decision-making context in Turkey, which have remained largely intact

for decades, as confirmed by the informants:

> The philosophy of education [in Turkey] is strictly based on the "official state ideology" in a very centralized educational system. (Turan, 2000, p. 554)

> Educational development in Turkey suffers from the lack of a holistic, evidence-based education strategy, one which builds on national consultation and consensus. (Aydagül, 2013, p. 226)

> Both primary and secondary education as well as higher education still remains geared toward constructing a monotype identity devoid of cultural pluralism found in Turkey's society, and maintains its ideological stance. (Çelik & Gür, 2013, p. 171)

This study reveals that education's top-down, centralized, and politicized nature

in Turkey (Keser-Aschenberger, 2012; İnal, 2012) has concealed the utility and

practicality of evaluation in the policy life cycle because of two complementary reasons.

First, extreme centralization paved the way to extreme bureaucratization that has

historically limited opportunities for collection and utilization of analytical evidence to

render judgments about decision making. Turan (2000) argued more than a decade ago

that the problem of a highly centralized system and resulting rigid bureaucracy still exists

in the Turkish education system, leaving little room for flexibility, commitment, and

ownership to policies, which is an obstacle against collective will for obtaining solid

feedback and fostering sustainable improvement. Preoccupation with practical formalities

and intensive clerical work under the guise of centralization, as John Dewey argued based

on his observations of the Turkish education system in 1924, paves the way to chronic

bureaucratization where intellectual properties – one of which is justifiably evaluative

thinking – are rarely if ever utilized. Dewey (as cited in Boydston, 1983) noted:

> There is also a danger that any centralized system will become bureaucratic, arbitrary and tyrannical in action, and given to useless and perfunctory mechanical work in making useless records, requiring and filling useless reports from others, and in general what is termed in French 'papasserie' and in English 'red-tape.' The functions of the Ministry should be intellectual and moral leadership and inspiration, rather than detailed administrative supervision and executive management. (pp. 280-281)

Second, the authoritarian prerogative of persistent centralization and

bureaucratization prescribes a symbolic institutional setup for knowledge systems to be

utilized that, in essence, undermines an intellectual and critical outlook in policymaking

by defining the borders and margins of that knowledge. Although institutionalization of

the evaluation function in the governmental arena has been considered a purposeful and

valuable act to systematically integrate evaluation into public decision making (see

Boyle, Lemaire, & Rist, 1999), its unintended consequences for the practice of evaluation

are not fully discussed in the literature. While discussing the benefits and drawbacks of

anchoring the evaluation function in different segments of government (i.e., legislative or

executive branches, corporate government unit), Mayne, Divorski, and Lemaire (1999)

posit that the location of evaluation significantly matters in defining the scope, purpose,

methodology, and utilization of evaluation studies. The authors argue that when

evaluation is anchored within a governmental organization too close to the program or

policy (i.e., the Ministry of National Education), evaluation activities tend to put more

emphasis on specifically assigned internal managerial, operational issues with limited

capability to deal with the technically and politically demanding issues of impact and

relevance. As such, evaluations' capacity to question the rationale and legitimacy of

certain programs or policies is significantly reduced.

Data from this study illustrate this point in the case of Turkey. The establishment of M&E units at MoNE thanks to Decree No. 652 was evidently a milestone for strengthening the value of evaluation as a decision-making tool. Yet, M&E units, both conceptually and practically, are believed to be located at the bottom of a highly structured institutional hierarchy, whose chances of visibly reorienting existing policies are contested, according to the study participants. The findings suggest that the units are invested in operational issues, dealing with managerial matters of performance, input-output balance, and project procedures, rather than overarching impact or the relevance of policies and programs that would otherwise challenge their continued existence (cf. Mayne, Divorski, & Lemaire, 1999). Additionally, drawing on Stufflebeam's (2002) 18-item checklist for institutionalizing the evaluation function, it is apparent that some elements of an effective institutional structure to sustain the value of evaluation are still missing in the Turkish educational decision domain, including but not limited to the nonexistence of a functional evaluation policy, standards and guiding principles for conducting and utilizing evaluations, effective communication channels between the M&E and other units, as well as ongoing evaluation training. As a result, coupled with a bureaucracy imbued with perceived uncritical clerical work, evaluation's institutional location in the Turkish educational domain may well further limit the visibility of the potential utility of evaluation.

Economic context. Second, the findings suggest that, despite the influence of educational politics to make opaque the value of evaluation for educational decisions, Turkey's economic context is pushing further to highlight the value of evaluation, hence

the challenging tension between the political and economic contexts for the evaluation enterprise.

The findings suggest that the linkages between economy and education are evident in Turkey. Turkey's quest for *development* and *modernization* and *reformation*, which are almost always repeated as the major goals for educational reform in the country, epitomizes the connection between economic growth and educational achievement promoted also by Turkey's development partners such as the World Bank, OECD, and the EU. The juxtaposition between economy and education holds that education is the most important investment that a country can make to sustainably increase economic production and eventually compete in the global-knowledge economy. Currently standing in 16th place, Turkey aspires to become one of the top ten largest economies in the world by the Republic's centennial year of 2023. The country's drive for economic development serves to further undergird the role of educational achievement and growth of an overwhelmingly young population, as discussed in Chapter 4 and confirmed by the study informants. To strengthen the connection between these two sectors, reform in educational governance has become necessary – facilitated by Turkey's development partners – that cultivates a culture of effectiveness, efficiency, and improved performance at the heart of which monitoring and evaluation has been located.

This observation is tied to the notion of *good governance* promoted by multilateral organizations such as the World Bank and OECD in order to spread a particular vision for policymaking, implementation, and public service delivery. Good governance envisions a sound public sector management grounded in the principles of

participation, accountability, and transparency by bringing state, civil society, and market closer to collaborate more effectively (Segone, 2008). As such, it is presented as a key ingredient for sustainable growth. With the advent of good governance, evidence-based or, more commonly, results-based management began to be associated with development as a means to ensure solid progress towards national development goals (UNDG, 2011). A results-based management (RBM) framework fosters a culture of decision making that aims to increase the effectiveness of government planning and programming by fully integrating monitoring and evaluation functions into the process so that results that matter are successfully achieved (UNDG, 2011). Many developed countries (e.g., the United States, Canada) quickly institutionalized RBM systems to embed accountability into public planning and budgeting (Lahey & Nielsen, 2013; Bourgeois & Cousins, 2013). Under the leadership of major international donors such as the UN, the World Bank, and OECD, numerous governments in less developed countries have also adopted RBM as a style of decision making and moved to centralize performance monitoring and evaluation functions in their public sector management. All in all, the connection between evaluation and development under the framework of good governance has been established and duly promoted.

Parallel to this, the trend of good governance has made its debut in Turkish public administration through educational projects funded by the country's international donors, as testified to by the informants. The influence of Turkey's international partners and donors in the country's educational outlook is one recurrent issue throughout the analysis of findings. The findings exposed that Turkey's ongoing reliance on supranational entities in providing conceptual, technical, and financial guidance for educational reforms

remains intact at this date. A recap of the historical background of these projects is necessary at this point. In the aftermath of World War II, the IMF and the World Bank were among the first comers into Turkey's education sector by providing credits and loans in the form of structural adjustment programs to reformulate the education system in accordance with the free market economy (Özdemir & Beltekin, 2012; Harvey, 2005). While earlier education projects targeted mostly vocational and technical training, the later projects especially the ones funded by the World Bank, were more comprehensive, aiming to restructure the entire education system (i.e., Basic Education Project I and II, Secondary Education Project). In the meantime, Turkey closely mimics OECD's education agenda as a member state by participating in global student tests such as PISA (Program for International Student Assessment) and TIMMS (Trends in Mathematics and Science Study), as well as by taking place in annual *Education at a Glance* reports (see OECD, 2013). Among all these actors, the European Union is the most influential player in reconfiguring and redirecting Turkey's education system in accordance with *Acquis*. The driving force behind a Europeanized educational policy is the desire to become the best in a competitive, global-knowledge economy (Rinne, 2008), and this is a mantra continuously repeated and promoted in almost all education reports authored and published by Turkish educational authorities.

Turkey's international partners have paved the way for a structural change in the education decision domain through these educational projects that have foregrounded unique concepts, including strategic planning, performance management, and evaluation in the country's public administration, as confirmed by the study participants and reinstated in policy documents (see also Eğinli, 2010; Çelik, 2012). The findings

demonstrated that these projects provided research and development – perceived to be of high quality – for Turkey's education sector that also repeatedly underlined the need for a change in educational governance. The EU's "*Strengthening the Capacity of the Ministry of National Education Project*" is the most influential project in this sense. Decree No. 652 that officially established the first M&E units in MoNE's history was conceptually grounded on the Green Paper published based on the project findings. Evidently, the passage of this decree was preceded by the passage of PFMC Act No. 5018 in 2003 that introduced strategic planning and performance-based budgeting into public management in the country. The document review suggested that Act No. 5018 was modeled after Turkey's development partners and justified on the grounds of strengthening accountability, transparency, effectiveness, and efficiency in public service delivery – the principles promoted with the advent of good governance as discussed earlier. These two laws are treated as testament to Turkey's surge of catching up with the developed world, further serving to establish the connection between evaluation – albeit under the disguise of such constructs as performance management and strategic planning – and economic development in the country.

In sum, the national context in Turkey, specifically the political and economic contours, has created a tension as far as the value of evaluation is concerned that results in locating evaluation in between full adoption and total abandonment. While Turkey's rapid economic development is working to strengthen the country's ties with development partners, mostly from the global North, and opening up a new era in public management where effectiveness and efficiency have become catch-phrases, inner

political contestations in the education sector are working to obscure the value of evaluation as a decision making tool to improve educational prospects.

   Contesting the concept of evaluation in Turkey and beyond. Given the scant to almost nonexistent empirical literature on the role and utility of evaluation in Turkey's public policymaking, much less the education sector, this study acknowledged the urgent need to explore how evaluation is conceptualized as a concept and practice in the country before discussing the need and capacity for conducting and utilizing evaluations for decision making in the first place. The study argued that how evaluation is conceptualized impacts the way it is valued in the public sector. At the onset, the study was grounded on the assumption that the concept of evaluation in the Turkish education decision making context will propose unique features that will challenge the current conceptions of the term recognized and utilized in other contexts. Overall, the findings indicate that this assumption was partially confirmed. This section will discuss how evaluation is still considered an emerging construct in the education sector, and how a lack of history with the concept and reported confusion about what the practice really entails are believed to contribute to the frequent association of evaluation with quantification. More importantly, the analysis of findings exposed an evolving tension regarding the meaning of evaluation. While reliance on quantifying educational processes and products was welcomed by some to clearly demonstrate Turkey's alignment with the developed world, others problematized this connotation by pointing to the neoliberal underpinnings of evaluation practice, and calling for a more culturally nuanced construction.

First and foremost, the findings reveal that evaluation is mostly associated with quantification and performance management. In a majority of the interviews (25 out of 35), evaluation was implicitly or explicitly discussed in the context of statistics, numbers, or student achievement test scores, as demonstrated in Chapter 5. Evaluation is mostly believed to provide quantified information regarding performance as well as compliance with rules and regulations by clearly exhibiting the connection between educational inputs and outputs. This notion explains why the most common espoused purpose of evaluation is to strengthen accountability in the Turkish education domain, as discussed in Chapter 5.

Two explanations were provided throughout the interviews as to why evaluation evokes a mental association with numbers. First, the findings suggest that evaluation remains as a new construct in the Turkish education decision domain, despite recent developments in the legal context and capacity building provided by development partners. A short history with the concept and practice is believed to preclude the development of professionalism in the field that would allow for more sophisticated constructions and configurations, yet currently evaluation is reduced to the level of methodology because it is the most convenient, and immediately accessible, as informants reckoned.

Second, some informants claimed that evaluation is a Western concept, working to expand the neoliberal agenda for educational policies, hence the common juxtaposition between statistics and evaluation. As discussed in the previous section, evaluation was officially put on the decision table through internationally funded projects. These informants critiqued the neoliberal undertone of these educational projects, specifically

problematizing the strong connection between economic production and educational growth. As a result, these respondents' preconceived notions of neoliberalism may have tainted their understanding of the concept and practice of evaluation. Their concerns are already established in the field of comparative education. In global education actors' reports, education is equated to the process of training the quality workforce responsive to the demands of global economy (Kumar & Hill, 2009; Grubb & Lazerson, 2004). It is argued that national education systems are thus reformulated to sustain economic development as prescribed by neoliberal ideology (Grant, 2009). Specifically, nation states are suggested to adhere to particular educational governance featuring performance-based budgeting and results-based management so that they can establish, strengthen, and maintain the principles of effectiveness, efficiency, and competitiveness, which were also repeatedly reported in the interviews and official documents for this study. According to these scholars, other inevitable features of this new form of governance that spill over to almost all public realms, including education, are standardization, quantification, categorization, and ranking that help reinforce commodification and marketization of the provision of education (İnal & Akkaymak, 2012; Kumar & Hill, 2009). According to some of this study's participants, it is this context that provides the clues for understanding why evaluation resonates with quantification and performance-based budgeting and even competitiveness.

Their observation can also be tied to the increased utilization of global student assessment studies such as *Education at a Glance*, PISA, TIMMS, TALIS, and PIRL in reassessing educational quality in the country (Steiner-Khamsi, 2003, 2004), as in the case of Turkey. Despite critics of the far-reaching policy implications of these test scores

(see Grisay & Griffin, 2006; Dolin & Krogh, 2010), many developing countries

(Argentina, Chile) use them as measuring sticks to compare themselves to other countries

(Valverde, 2014). Similarly, in the absence of systematic, formal, national policy

evaluations in Turkey coupled with the country's long-term desire for modernization,

these global achievement tests are frequently used to "evaluate" the outcomes of

educational policies and projects in Turkey (see Aydagül, 2008; EARGED 2005, 2006).

An illustration of this can be found in Çelik and Gür's (2013) analysis of the JDP

government's educational policies. The authors noted:

> In order to correctly determine the improvement in the quality of education following a series of reforms, it would help to assess the effectiveness and efficiency of government policies as well as the remaining weaknesses of Turkey's education system. International studies including TIMMS and PISA provide valuable data for such analysis. (p. 156)

Parallel to this, one of the leading nongovernmental actors in shaping economic

outlook in Turkey, namely the Economic Policy Research Foundation (TEPAV),

endeavored to draw a linear line between PISA scores and the industrialization of Turkey.

Problematizing Turkish students' standing in PISA rankings with special reference to

their average math and science scores, the Foundation made projections about the welfare

of Turkish citizens. Specifically, the authors noted, "The achievement level of students

who have not yet stepped into [the] job market will determine the capabilities of future

labor force as well as their capacity for innovation, hence production model's

technological structure in the future" (TEPAV, 2008, p. 6). Taking the developed

countries as a benchmark, the foundation recommended that Turkish educational

authorities dwell into the reasons of falling short of criteria for educational excellence

demonstrated by the PISA scores and take advantage of the demographic advantage

offered by a young population. Nevertheless, the statistical information and comparative

analysis in the form of league tables provided by these global tests and reports (Steiner-Khamsi, 2003) further encourage the exclusive association of evaluation with sheer testing and numbers, missing the larger level of decision making that is much more nuanced and complicated than imagined and recommended (Steiner, March 2014). This is consistent with the current conceptualizations of evaluation in the Turkish education domain where evaluation is conceived as a tool for quantification and compliance to ensure – down the road – Turkey's alignment with the global knowledge economy.

A tension emerged at this juncture whereby some respondents simply welcomed ongoing reliance on quantitative data to demonstrate Turkey's level of development and alignment with the global education standards, while some argued for its limited contribution to educational decision making. Nevertheless, a consensus emerged that a contextualized construction of the concept of evaluation is still needed. The findings indicate that concepts like evaluation are socially, politically, and culturally constructed to remedy issues that are unique to the context in which they were born. Thus, the evaluation concept will justifiably be contested, resisted, and negotiated in different settings, as evidenced in the case of Turkey.

The request for localizing the concept to better serve the educational problems in Turkey in a more responsive manner calls for bringing back the literature on emerging indigenous evaluation practice in the developing country context, discussed in Chapter 2. There has been recent theorizing about decolonizing evaluation practice in our field, addressing such critical questions as: Whose agenda is being served by development evaluation sponsored by Northern donors? Whose values and ideas does evaluation signify in a national context? Whose questions are answered to what purpose? How are

findings disseminated and utilized for whom? (see Hopson, Kirkhart, & Bledsoe, 2012; Mertens, 2008; LaFrance & Nichols, 2008). A recent forum in the *American Journal of Evaluation* (AJE) on the international growth of the evaluation profession revitalized the indigenous evaluation practice (Schwandt et al., 2013) that was initially put forward by Carden (2010) in another forum of *AJE*. As summarized by the director of the United Nations Development Group in the forum, "There is a need for greater attention to ensure that evaluation is an indigenous process, rather than a thrust from outside, and there are many successful examples, where national governments took a lead in establishing evaluation functions" (p. 574). Although the desire to indigenize the evaluation concept in the Turkish education domain, aligned with Turkey's perceived material, moral, and ethical values – hospitality, self-sacrifice, compassion, patriotism, and heroism (Inal, 2012) – how this can be achieved is unknown, and the study findings could not provide further insight into this issue.

    Revisiting the need and capacity for evaluation in Turkey and beyond. After establishing the contextual and conceptual parameters of evaluation, this study aimed to explore the need for evaluations and the existing capacity to conduct and use evaluations in the Turkish educational decision domain. The study reasons that the existing demand for and supply of evaluation will impact the value ascribed to the evaluation function. Absolutely none of the respondents interviewed for this study doubted the need to conduct and utilize evaluations to obtain systematic feedback about policies. They identified certain elements in place that pave the way to that expectation, such as centralized database management, educated staff members, and a legal mandate for evaluations. Yet, the findings suggested a subtle hierarchy among the two significant

elements of evaluation capacity for the Turkish case: the legal environment, and human capital. First, the study identified reservations about the legal enabling environment for evaluation, which is commonly noted in the literature among the most important factors to establish the value of evaluation for decision making in the developing world. Second, the study spotlighted the important role played by evaluation champions in advancing an evaluation agenda in a context where it is historically immature and put forward the argument that it is the human capital that *enables* the environment (either legal or political) for evaluation to flourish.

First and foremost, the findings suggest that the legal enabling environment for building the value of evaluation in the Turkish education domain has been provided by Decree No. 652. Despite the respondents' visible excitement about the passage of this piece of legislation, the majority was quick to mention that the legal context is necessary but not sufficient to maintain the evaluation enterprise in the Turkish context. They posit that the decree calls for a change in culture and a change in thinking about how programs and policies are designed, implemented, and modified, and the findings exposed that it will take considerable time to establish such a mentality. Uniting these comments was a common acknowledgment of the role of individuals in appreciating and making meaning of this change as well as in materializing this demanded change in mindset. To sum, it is contended that the legal environment cannot be readily translated into sustained valuing of the evaluation function.

This last finding brings us to the second, and seemingly more important, point about the existing human capital for building and sustaining the value of evaluation that presents a challenging tension in the case of Turkey. On the one hand, the study

participants praise the new, young recruits at the MoNE and expect that their energy and enthusiasm as well their level of education and perceived innovativeness will fuel effective educational planning and programming and foster the value of evaluation. On the other hand, civil service mentality is commonly cited as an inherent obstacle to these young bureaucrats' desire to think critically and expand their perspectives to investigate the shortcomings of policies. In addition, the findings exposed some reservations about the new recruits' primary source of evaluation knowledge and their level of technical evaluation skills – given that the field of practice has not been fully developed in the country. A need for a more solid understanding and development of evaluation knowledge, skills, and ability especially among M&E staff at MoNE has been warranted to build and sustain the role and utility of evaluation.

All in all, this study identified the existence of evaluation champions as the main drivers to promote and cement the value of evaluation in the education decision domain. The vitality of evaluation champions to growing an evaluation culture at the mesa-organizational level is commonly cited in the mainstream literature (e.g., Cousins & Bourgeois, 2014; King & Volkov, 2005; Taylor-Powell & Boyd, 2008). Patton (2008) markedly highlight the significance of champions, building his Utilization-Focused Evaluation (UFE) based on the *personal factor*, defined as an individual or a group of individuals who care about evaluations and assume a personal responsibility to use results. These caring individuals in the Turkish context have worked to present evaluation as a tool for decision making. In this study, the former Minister of National Education, Dinçer, was repeatedly mentioned as the brain child of Decree No. 652, hence the pioneer in pushing forward the evaluation agenda in the education sector. His educational and

professional background as an academic and his perceived valuing of evidence are thought to foreground this agenda. In addition to such a strong individual actor, civil society, more especially educational think-tank organizations, emerged as influential advocates for improving educational planning and policymaking based on solid and systematic information obtained from evaluation. A sense of hope is developing among the participants that the existence of evaluation champions will increasingly serve the need for increased governmental responsiveness to societal needs.

In sum, the study findings underscored the role of individuals in discovering, highlighting, building, and sustaining the value of evaluation in a setting where evaluation is still emerging as a form of inquiry. In doing so, the study aims to change the focus of the conversation from the supply-oriented, technical aspects of evaluation capacity such as database management systems, financial assets, legal mandates (Rist, Boily, & Martin, 2011) – which the Turkish education domain is believed to already encompass to a great extent – more into the demand side whereby individuals – better yet champions – make meaning and use of evaluations.

Current status of evaluation in the Turkish education domain. Drawing on the context for, the concept of, and the need and capacity for evaluation in the Turkish education decision domain, a picture has emerged regarding the current utilization of evaluation findings and processes. Despite a legal mandate for evaluation practice, as well as integrated, full-fletched database management systems and educated staff members, existing evaluation practice does not reflect a stable character in the education sector. The reasons for this observation are many.

First, evaluation is still a mystery, despite the built-up expectations if the deputy ministers or even rank-and-file bureaucrats are being held accountable for the performance of the programs and policies. Second, the absence of well-defined, formal structures and procedures to conduct and utilize evaluations speaks to the ascribed value of evaluation since evaluations are likely to be conducted on an *ad hoc* basis under the current framework, being more susceptible to fluctuations in the political landscape, as feared by the study informants. Third, the Ministry of Development's reported struggle with conducting and utilizing evaluations across governmental agencies tells much about the struggle MoNE has to go through, given the hierarchy of decision making. When the evaluation enterprise is not yet fully incorporated into national development plans at the top of the hierarchy, the findings suggest that ministerial adoption will be partial as well. Last but not least, the introduction of M&E units facilitated a systematic *monitoring* through the centralized database management systems at best, instead of evaluating the relevance or effectiveness of large policy initiatives in the domain.

Under these circumstances, the role and utility of evaluation is mostly materialized and recognized through ex-ante pilot studies based on which the system-wide adoption of educational initiatives is legitimated. In the meantime, international donors continue to fund external evaluations of their educational projects, and domestic think-tank organizations endeavor to inform the public of the outcomes of policies based on external research (see ERI, 2013a, 2013b).

Concluding remarks. To return to the overarching question this study addressed – what is the value of evaluation as a decision-making tool in the Turkish educational decision domain? – the integration of findings shows that evaluation's value resides in its

promise to help Turkey reach the level of most developed countries of the global North. Turkey's development is tied to the extent to which she adheres to the new educational governance at the heart of which monitoring and evaluation reside. As such, evaluation has been a benchmark, a measure to be taken seriously in the face of an ever changing global economy. Evaluation is a strong indicator of compliance to international best practices, keeping Turkey in line and moving forward with global trends. Evaluation is a testament to Turkey's resilient surge of and continued commitment to modernization.

While evaluation is visibly entertained on the foreign front, its domestic reflections are elusive due to its political nature. Hence, evaluation is in foster care within the borders of Turkey, specifically in the education sector. The physical and mental space evaluation occupies in Turkish educational decision making is but a shadow, lacking serious embodiment of built-up expectations resulting from ongoing administrative improvements and developmental promises. Although evaluation is recognized as a familiar, valuable tool, the swinging political climate keeps evaluation's actual value for educational decision making in limbo. The evaluation concept has been tainted by actors' preconceived and normative understandings of Westernization, development, modernization, and even neoliberalism; hence, it is suggested that evaluation would constitute a larger significance in governmental life if Turkish actors had a way to indigenize it for their own information needs. At this very juncture, individuals' role emerged as a significant enabler for building the value of evaluation in Turkey.

Recommendations for Research and Practice

Based on the discussion of findings presented above, the purpose of this section is to explicitly detail how this study advances the existing literature by highlighting the

remaining gaps in our knowledge about the concept and practice of evaluation. Similar to the previous one, this section will explicate the study's contributions to research and practice as well as remaining gaps in accordance with the research questions explored, namely, the context for evaluation, the concept of evaluation, and the need and capacity for evaluation. Recommendations are meant to inform evaluation research and practice simultaneously as this study does not assume any clear, meaningful distinction between the two as far as the field of evaluation is concerned.

On the context for evaluation. Over the past few years, context – cultural, social, political, and historical – has dominated the conversation on evaluation's use and influence. Despite a plethora of studies that point to the significance of context in defining the terms of evaluation in organizational settings (Cousins & Bourgeois, 2014), empirical accounts of why and how evaluation as a form of inquiry is negotiated and altered in a low or middle income country's governmental decision-making setting are still rare. A comprehensive issue of *New Directions for Evaluation* titled "Context Matters" edited by Conner, Fitzpatrick, and Rog (2012), has been the closest attempt in – borrowing the authors' own words – "bringing context from background to foreground" in decision-making settings (Rog, 2012, p. 26). Yet, theirs still did not unpack – nor did it have the intention to – the specific elements in a middle-income, non-Northern country context whose history with evaluation practice does not go back far. This is the gap these findings endeavored to close, drawing on the country case of Turkey. Specifically this study advances the literature on the context for evaluation in two significant ways. One concerns the juxtaposition of political structures and evaluation, and the other takes into account the field of practice in which evaluation can be established as a valued enterprise.

First and foremost, the study findings point out the issue of what specific role evaluation can assume in a context where education has traditionally been conceptualized and materialized as a tool of control and indoctrination of state ideology. While evaluation's incorporation into decision making is believed to mirror participatory democracy in the global Northern context (Green, 2005), drawing on Çelik and Gür's (2013) remarks about the Turkish "education system's excessively centralized organizational structure as well as the presence of overly nationalistic expressions in school curricula that leaves no room for pluralism and seeks to indoctrinate the student body" (p. 172), it is safe to argue that evaluation's role to promote a diversity of voices, viewpoints, and realities is likely to be significantly limited in Turkey. When the governmental decision agenda seeks to promote and disseminate particular values in a society by default, the following questions should merit attention in moving forward with expanding this field of practice into virgin contexts: What value does evaluation ultimately seek to address in a particular public policy area? What is the ethical warrant for evaluative judgments when the states establish evaluation systems to aid a particular vision for decision making? Can evaluation's value where state dogmatism exists ever be legitimated?

Second, scholars have emphasized the trans-disciplinary nature of evaluation (Scriven, 1994) *as though* evaluative thinking and the installment of evaluation systems could take roots and prosper in any field of practice equally. This assumption grossly treats the governmental decision-making context as a holistic entity – easy to study, comprehend, and manipulate – whereas experience clearly suggests that it is actually larger than the sum of its parts, including different line ministries with different policy

areas. Depending on organizational structure, staffing requirements, and engagement with domestic and international counterparts and, more importantly, the degree of politicization and polarization in its area of jurisdiction, a governmental body's likelihood of internalizing evaluative thinking and activities will be immensely different. For example, while the field of education has been a hotbed for evaluative thinking to flourish and expand in many developed countries (e.g., the United States, Canada, and the United Kingdom), given the significance and politicization of educational services for national development and security, the field of education in Turkey may not be a convenient and effective choice for evaluation champions to build systems and practice. Hence, the following questions still beg for an answer for which this study provided the leeway: Which fields of practice – or the areas of governmental jurisdiction – are much more welcoming to building and maintaining the value of evaluation in a country context? How and why does it make a difference to introduce evaluation through a particular sector?

On the concept of evaluation. Multiple and differentiated meanings of evaluation as a form of inquiry in the Turkish education sector revealed that the discussion on the philosophical foundations of the concept of evaluation in the global South is immature in existing empirical and anecdotal accounts. Existing literature treats evaluation as an exclusive, single-handed phenomenon that essentially exists in a vacuum and shapes the direction and nature of decisions almost in a unidirectional fashion. The findings of this study indeed suggest that paradigmatic foundations of any evaluation system will be grounded and justified on the philosophical underpinnings of policies that the very evaluation system is established to address (for an extensive discussion of

distinct educational philosophies and their reflections on policies, see Turan, 2008). As a result, this study advances the literature on the concept of evaluation in two unique ways: the first one relates to the philosophical compatibility between policy agendas and evaluation regimes that may not seem apparent at first sight, as in the case of Turkey; and the second one, building on the first point, draws attention to the difficulties in creating an indigenous construction of evaluation in an era where a particular and a relatively dominant vision for evaluation may already have been constructed.

To begin with, the study findings on the concept of evaluation alert us to the question of why certain conceptions of evaluation (i.e., quantification, performance-based budgeting, and compliance as in the case of Turkey) supersede others in a given policy context. The argument was made earlier that the kinds of educational policies that are being promoted in Turkey explain the kinds of evaluation concepts and systems envisioned and structured. The pursuit of modernization and development provide the rationale for Turkey's current educational outlook. To this end, as pointed out in the previous section, the success of Turkey's educational system is nowadays assessed by the rate of return in the job market, and this connection is facilitated by international student achievement tests. This signals a need for a particular evaluation regime favoring the ubiquitous discourse on quantification in the development landscape. Thus, this study invites evaluation scholars and practitioners to attempt to answer the following questions: How and in what ways does the ideological orientation of government policies (education, health, environmental, etc.) necessitate a particular vision for evaluation? How does globalization of national education policies impinge on the concept of evaluation in LMICs?

At this point, this study aims to redirect scholars' attention to the role of multilateral and bilateral donors in potentially standardizing the value of evaluation for decision making in a developing country context. Internationalization and globalization are believed to be the major driving forces behind reform agendas across countries (Boli, Ramirez, & Meyer, 1985; Philips, 2004; Çelik, 2012), orchestrated by supranational actors such as OECD, World Bank, IMF, EU, UNICEF, and UNDP, as demonstrated in the case of Turkey. These powerful actors seek to determine the direction and content of national educational policies around the world through standards and criteria for educational excellence (Steiner-Khamsi, 2004). Educational studies, reports, statistics, and test scores administered, conducted, compiled, and published by these organizations may be, at times, treated as the *truth* in education sector, demarcating a zone of authority that members are expected to submit (Rinne, 2008). Therefore, caution is warranted because, while scholars in the field of evaluation have been pushing for more diversity in evaluation approaches, systems, and practice from design to utilization, supranational actors may be pulling towards a more homogenized evaluation regime across the world driven primarily by methodological concerns. Evaluation scholars and practitioners cannot afford to overlook the terms and conditions of evaluation that developing countries are pressured to accept. As a result, a daring question of whether evaluation systems are reinforcing colonizing assumptions of power and authority by their potential discourse of standard setting should merit attention in the field of evaluation (Steiner-Khamsi, March 2014).

This last point brings us back to one of the earlier arguments of this study about the indigenous evaluation practice in the global South. While the need for ceding control

over evaluation agendas from donors to the developing country people is starting to be discussed in the literature (Carden, 2007), the question of how a locally-relevant evaluation mindset can be established is still unclear. One the one hand, this study points out that the notion that evaluation plays a pivotal role in every governmental decision-making process – uninhibited by political systems, bureaucratic cultures, and historical structures – runs the risk of essentializing the concept of evaluation and practice to an extent that the meaning, margins, and negotiations of evaluation's value in unadulterated contexts are overlooked. In *Decolonizing Methodologies*, Smith (2012) pointed to this essentialist tone in Western practices of research and its cognate field of evaluation:

> Research "through imperial eyes" describes an approach which assumes that Western ideas about the most fundamental things are the only ideas possible to hold, certainly the only rational ideas, and the only ideas which can make sense of the world, of reality, of social life, and of human beings. It is an approach to indigenous peoples which still conveys a sense of innate superiority and an overabundance of desire to bring progress into the lives of indigenous peoples – spiritually, intellectually, socially, and economically. (p. 58)

Parallel to Smith's argument, this study also invites us to be alerted to ethnocentric assumptions of evaluation practice that may be established and routinized over time by institutional and governing structures and duly transferred to indigenous contexts.

On the other hand, this study did not find any empirical opportunity to further unpack what an "indigenously" or "authentically" Turkish evaluation practice might look like, other than exposing that there is a need for one. To be specific, the respondents clearly mentioned that the Turkish culture would provide a solid foundation for building a local mindset for the practice, yet the guidelines to achieve an authentic Turkish evaluation practice were not discussed. The field of evaluation, in the meantime, has been slow to engage in theoretical and philosophical conversations about indigenous evaluation practice in contexts outside of the global North (for an exception see Carden &

Alkin, 2012). Hence, this study seeks to rejuvenate the theoretical conversation by calling scholars to investigate the following questions: What does the term "indigenous" mean in a given country context? What do an indigenous evaluation concept and practice look like in a given country context? What are the implications of indigenous evaluation for the field of evaluation beyond a conversation on methodological choices?

On the need and capacity for evaluation. ECB in a developing country context has been advanced and explored in various books and studies (e.g., Boyle & Lemaire, 1999; Segone, 2008, 2009; Tarsilla, 2012), and this study's findings contribute to this bourgeoning literature in two significant ways. First, this study partially – and indirectly – provides empirical evidence for the success of ECB activities designed and organized by the aid organizations in the global South. Second, the study puts forward an argument that there might be a hierarchical relationship between the elements of evaluation capacity identified in the literature.

To begin with, at the outset, this study problematized the rationale for and the success of evaluation capacity building (ECB) in the global South, orchestrated primarily by Northern-based aid organizations. It was extensively discussed in Chapter 2 that these efforts may serve the needs of donors for a sustainable assessment of the impact of their development projects more so than the needs of developing country partners. Given the study's conclusion that evaluation still remains an emerging construct in the Turkish education decision domain (Türk, Yalçın, & Ünsal, 2006) despite the plethora of educational projects funded by Turkey's development partners, as well as ECB activities provided for the education sector (Stout, 2010), the donors' success in streamlining evaluation in public sector management in LMICs is justifiably contested. More

especially, the suggested relationship between a sense of ownership of policies and the need for evaluations in a national context is further telling about the success of ECB efforts. Some participants argued that a need for policy evaluations, underneath, drives from a sense of ownership to those policies, which is, in large measure, distorted by the continuous involvement of international actors in the Turkish education sector. It was probably this notion that partially caused the critics to speak of the education sector as a "jigsaw puzzle" or a "project dumpster." This argument counsels a critical caution: While donors invest in ECB, working to advance the evaluation agenda, their policy initiatives may reduce local actors' feeling of ownership to those policies that in return discredit the need to evaluate. Hence, this study invites scholars and practitioners alike to be critical of the success of ECB efforts in the global South, investigating the unintended consequences within the larger context of development initiatives.

Second, and parallel to the previous point, the study was premised on the argument that ECB is context specific (King, 2007), hence the warrant of contextual variations of a broader set of elements related to evaluation capacity. Although ECB at the meso (organizational) level was not the primary focus of this study – the attention was rather directed towards the macro level of the education domain incorporating more than the Ministry of National Education as the focal point – these findings may provide an opportunity for a more critical outlook for understanding and differentiating ECB efforts across domains, sectors, or organizations. By putting the individuals at the center of ECB conceptual or practical frameworks and guidelines, the Turkish case suggests that there is perhaps an order to building capacity in a context where evaluative thinking or culture is historically underdeveloped. The existing literature on ECB treats all elements

of evaluation capacity equally important, implying that all need to be eventually in place for evaluation to be cemented in an organization (e.g., Nielsen, Lemire, & Skov, 2011; Stockdill, Baizerman, & Compton, 2002). This study, on the contrary, requests a systematic, empirical investigation of the specific role of evaluation champions in a developing country context in building a valued evaluation enterprise in the public sector guided by the following questions: Who is an evaluation champion in a developing country context? How does one become a champion in a context where an evaluation culture is historically missing?

To sum up, drawing on social science literature from the fields of evaluation, comparative education, public policy, and international development, this study revisited the origins of evaluation practice in the global Northern context, traced its expansion into the global South across a number of sites, and argued that context matters in transferring, borrowing, negotiating, establishing, practicing, and using the concept and practice of evaluation. By suggesting that evaluation is a marker of a country's quest to modernize and Westernize, this study sheds light on the direction of cross-cultural expansion of the field of evaluation.

Epilogue

Turkey's historic image as a bridge connecting East to West, fusing the global South and North in its unique landscape, will hardly, if ever, get old. A year does not pass by without domestic or foreign players finding a way to demonstrate different derivatives of this image, and this study is only one of them. As a country that has long set one eye on the West, evaluation's potentially fleeting status in Turkish governmental life upsets the global knowledge economy's authority and logic. Since the advent of good governance, a particular structure for educational governance has been promoted for the sake of efficiency and effectiveness, and Turkey has been called on to commence a series of political and administrative reforms. Establishment of monitoring and evaluation units at the Ministry of National Education was a symbolic, strategic response to that call. However, the country's established political structures and cultural outlook justifiably put evaluation under a dimmed light, mirroring the paradoxical nature of development the country has comfortably fallen into. Turkey wants to eat her cake and have it, too.

Evaluation's value is bounded by time and space in Turkey, significantly mediated by the political ideologies of its time and the modernizing assumptions it advances, hence it remains unattended at worst and controlled at best. Turkey's strategic refrainment from fully materializing the value of evaluation for governmental decision making invites evaluation scholars and practitioners to wrestle with normative questions of development, modernization, and decolonization and to assemble to bear evidence with theoretical and empirical rigor on discussions surrounding the cross-cultural expansion of our practice.

References

Adrien, M. H., & Jobin, D. (2009). Country-led impact evaluation: A survey of development practitioners. In M. Segone (Eds.), *Country-led monitoring and evaluation systems: Better evidence, better policies, better development results* (pp. 102-118). New York, NY: UNICEF.

Alkin, M. C. (1969). Evaluation theory development. *Evaluation Comment*, *2*(1), 2-7.

Alkin, M., & Taut, S. (2003). Unbundling evaluation use. *Studies in Educational Evaluation*, *29*(1), 1-12.

Alkin, M. C. (2004a). Context-adapted utilization: A personal journey. In M. C. Alkin (Eds.), *Evaluation roots: Tracing theorists' views and influences* (pp. 293-303). Thousand Oaks, CA: Sage Publications, Inc.

Alkin, M. C. (2004b). *Evaluation roots: Tracing theorists' views and influences*. Thousand Oaks, CA: Sage Publications, Inc.

Alkin, M. C. (Ed.). (2012). *Evaluation roots: A wider perspective of theorists' views and influences*. Thousand Oaks, California: Sage Publications, Inc.

Almeida, C., & Báscolo, E. (2006). Use of research results in policy decision-making, formulation, and implementation: a review of the literature. *Cadernos de Saúde Pública*, *22*, 7-19.

American Evaluation Association. (2004). Guiding principles for evaluators. Retrieved from http://www.eval.org/Publications/GuidingPrinciples.asp

Anderson, J. E. (2006). *Public policy making* (6th ed.). Boston: Houghton Mifflin Company.

Aslan, G., Küçüker, E., & Adıgüzel, E. (2012). External education projects in Turkey. In K. İnal & G. Akkaymak (Eds.), *Neoliberal transformations of education in Turkey* (pp. 95-107). New York: Palgrave Macmillan.

Atwood, J. B. (2012). Development co-operation report 2012. Retrieved from http://www.oecd-library.org/docserver/download/4312011e.pdf?expires=1353785529&id=id&accname=guest&checksum=BCDC4B77C28FC2853B4886B413CFD209

Aydagül, B. (2008). No shared vision for achieving Education for All: Turkey at risk. *Prospects*, 38(3), 401-107.

Aydagül, B. (2013). Turkey: Translating new policy paradigms to results. In Y. Wang, (Eds.), *Education policy reform trends in G20 members* (pp. 207-227). Berlin, Heidelberg: Springer-Verlag.

Aydın, A. (1996). *Milli eğitim politikaları ve şuralar* [National educational policies and councils]. Ankara: Ministry of National Education.

Ba Tall, O. K. (2009). The role of national, regional, and international evaluation organizations in strengthening country-led monitoring and evaluation systems. In M. Segone (Eds.), *Country-led monitoring and evaluation systems: Better evidence, better policies, better development results* (pp. 119-134). New York, NY: UNICEF.

Ba Tall, O. K. (2011). Improving evaluation demand to foster accountability and public management. In United Nations Development Programme (UNDP) (Eds.), *National evaluation capacities. Proceedings from the international conference* (pp. 136-148). Washington, D.C.: Evaluation Office.

Babacan, A., & Serdengeçti, S. (2003, April). Turkey – Letter of Intent. Retrieved from http://www.imf.org/External/NP/LOI/2003/tur/01/

Balthasar, A., & Rieder, S. (2000). Learning from evaluations effects of the evaluation of the Swiss Energy 2000 programme. *Evaluation*, *6*(3), 245-260.

Bamberger, M. (1991). The politics of evaluation in developing countries. *Evaluation and Program Planning*, *14* (4), 325-339.

Bamberger, M., Rugh, J., & Mabry, L. (2012). *RealWorld evaluation: Working under budget, time, data and political constraints*. Thousand Oaks, CA: Sage Publications, Inc.

Barkçın, F. (2002). *Avrupa Birliği'ne giriş süreci içinde VII. ve VIII. beş yıllık kalkınma planlarında eğitim politikaları* [Educational policies in the 7th and the 8th 5-year development plans in the process of entering the European Union]. Istanbul: Alman Liseliler Kültür ve Eğitim Vakfı.

Bhola, H. S. (2003). Social and cultural contexts of educational evaluation: A global perspective. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 397-414). Dordrecht, Netherlands: Kluwer.

Birkland, T. A. (2005). *An introduction to policy process: Theories, concepts, and models of public policy making* (2nd ed.). New York: M.E. Sharp.

Bishop, R. (2012). Indigenous methods in qualitative educational research. *Handbook of Qualitative Research in Education*, 126.

Boli, J., Ramirez, F. O., & Meyer, J. W. (1985). Explaining the origins and expansion of mass education. *Comparative Education Review*, *29* (2), 145-170.

Booth, D., & Lucas, H. (2001). *Desk study of good practice in the development of PRSP indicators and monitoring systems: Initial review of PRSP documentation*. London: Overseas Development Institute.

Bourgeois, I., & Cousins, J. B. (2013). Understanding dimensions of organizational capacity building. *American Journal of Evaluation*, 34 (3), 299-319.

Boydston, J. A. (Eds.). (1983). *John Dewey: The middle works, 1899-1924*. Carbondale, IL: Southern Illinois University Press.

Boyle, R., & Lemaire, D. (Eds.). (1999). *Building effective evaluation capacity: Lessons from practice*. New Brunswick, NJ: Transaction Publishers.

Boyle, R., Lemaire, D., & Rist, R. C. (1999). Introduction: Building evaluation capacity. In R. Boyle, & D. Lemaire (Eds.), *Building Effective Evaluation Capacity* (pp. 1-22). New Brunswick, NJ: Transaction Publishers.

Carden, F. (2007, December 1). The real evaluation gap. *Alliance Magazine*, *12*, 53-54.

Carden, F. (2010). Introduction to the forum on evaluation field building in South Asia. *American Journal of Evaluation*, *31*(2), 219-221.

Carden, F., & Alkin, M. C. (2012). Evaluation roots: An international perspective. *Journal of Multi Disciplinary Evaluation*, *8* (17), 102-118.

Çelik, Z. (2012). *The transformations experimented within the Turkish education system in the context of policy and implementation: The case of 2004 curriculum reform.* (Unpublished doctoral dissertation). Hacettepe University, Ankara.

Çelik, Z., & Gür, B. S. (2013). Turkey's education policy during the AK Party era (2002-2013). *Insight Turkey,* 15(4), 151-176.

Chelimsky, E., & Shadish, W. R. (Eds.). (1997). *Evaluation for the 21st century: A handbook*. Thousand Oaks, CA: Sage Publications, Inc.

Chelimsky, E. (2006). The purposes of evaluation in a democratic society. In I. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The Sage handbook of evaluation* (pp. 33-55). Thousand Oaks, CA: Sage Publications, Inc.

Chouinard, J. A., & Cousins, J. B. (2009). A review and synthesis of current research on cross-cultural evaluation. *American Journal of Evaluation*, *30*(4), 457-494.

Christie, C. A. (2007). Reported influence of evaluation data on decision makers' actions an empirical examination. *American Journal of Evaluation*, *28*(1), 8-25.

CIA Factbook. (2013). The world factbook: Turkey. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/geos/tu.html

Clarke, A. (2008). Evidence-based evaluation in different professional domains: Similarities, differences and challenges. In I. Shaw, J. C. Greene, & M. M. Mark (Eds.). *Handbook of evaluation: Policies, programs and practices* (pp.559-581). California: Sage Publications, Inc.

Conlin, S., & Stirrat, R. (2008). Current challenges in development evaluation. *Evaluation*, *14*, 193-208.

Conner, R. F., Fitzpatrick, J. L., & Rog, D. J. (2012). A first step forward: Context assessment. In D. J. Rog, J. L. Fitzpatrick, & R. F. Conner (Eds.), *Context: A framework for its influence on evaluation practice*. New Directions for Evaluation, *135*, 89–105.

Contandriopoulos, D., & Brousselle, A. (2012). Evaluation models and evaluation use. *Evaluation*, *18* (1), 61-77.

Cook, T. (1997). Lessons learned in evaluation over the past 25 years. In E. Chelimsky, & W. R. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 30-52). Thousand Oaks, CA: Sage Publications, Inc.

Cooper, M. (2002). The legacy of Atatürk: Turkish political structures and policy-making. *International Affairs*, 78 (1), 115-128.

Cousins, J.B., Goh, S., Clark, S., & Lee, L. (2004). Integrating evaluative inquiry into the organizational culture: A review and synthesis of the knowledge base. *Canadian Journal of Program Evaluation*, 19 (2), 99-141.

Cousins, J. B., & Bourgeois, I. (Eds.). (2014). Organizational capacity to do and use evaluation. *New Directions for Evaluation*, 141.

Court, J., & Young, J. (2003). *Bridging research and policy: Insights from 50 case studies*. London: Overseas Development Institute.

Court, J., & Young, J. (2004). Research impact on policy: What can researchers do? *Newsletter of the Economic Research Forum for the Arab Countries, Iran & Turkey*, 11(1), 18-25.

Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches*. Thousand Oaks, CA: Sage Publications.

Creswell, J. W., & Clark, V. L. P. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage Publications.

Dabelstein, N. (2003). Evaluation capacity development: Lessons learned. *Evaluation*, 9(3), 365-369.

Dahler-Larsen, P., & Schwandt, T. A. (2012). Political culture as context for evaluation. In D. J. Rog, J. L. Fitzpatrick, & R. F. Conner (Eds.), *Context: A framework for its influence on evaluation practice*. New Directions for Evaluation, 135, 75–87.

Datta, L. E. (2011). Politics and evaluation: More than methodology. *American Journal of Evaluation*, *32*(2), 273-294.

Davutoğlu, A. (2008). Turkey's foreign policy vision: An assessment of 2007. *Insight Turkey*. 10, 77-96.

Deniz, M. (2001). *Milli eğitim şuralarının tarihçesi ve eğitim politikalarına etkileri* [National education councils' history and impact on educational policies]. (Unpublished master's thesis). Anadolu University, Eskisehir.

Derviş, K., & Serdengeçti, S. (2002, July). Turkey- Letter of Intent. Retrieved from https://www.imf.org/external/np/loi/2002/tur/04/index.htm

Dinçer, O., & Yılmaz, C. (2003). *Kamu yönetiminde yeniden yapılanma 1: Değişimin yönetimi için yönetimde değişim* [Restructuring in public management 1: Management of change for change in management]. Ankara: The Presidency of the Republic of Turkey.

Donaldson, S. I. (2008). In search of the blueprint for an evidence-based global society. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds), *What counts as credible evidence in applied research and evaluation practice?* (pp. 2-18). Thousand Oaks, California: Sage.

Dolin, J., & Krogh, L. B. (2010). The relevance and consequences of PISA science in a Danish context. *International Journal of Science and Mathematics,* 8(3), 565-592.

Dye, T. R. (2008). *Understanding public policy* (12th ed.). New Jersey: Prentice-Hall.

EARGED. (1997). *Milli Eğitim Bakanlığı'nda program çalışmaları* [Curriculum studies at the Ministry of National Education]. Ankara: Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Başkanlığı.

EARGED. (1999). *Müfredat laboratuvar okulları gelişimi değerlendirme araştırması* [Evaluation of the development of the curriculum laboratory school]. Ankara: Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Başkanlığı.

EARGED. (2002). *OBBS 2002 ilköğretim öğrencilerinin başarılarının belirlenmesi ve durum belirleme raporu* [Determining the achievement level of primary school students in OBBS 2002 and report on the current situation]. Ankara: Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Başkanlığı.

EARGED. (2005). *PISA 2003 projesi ulusal nihai rapor* [Final national report on PISA 2003 study]. Ankara: Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Başkanlığı.

EARGED. (2006). *Temel eğitime destek programı öğretim programlarının (6. sınıflar Türkçe, Matematik, Fen ve Teknoloji, Sosyal Bilgiler) 1. Değerlendirme raporu: Matematik dersi* [The first evaluation report of the support to basic education program curriculum (6th grade Turkish, Math, Science and Technology, and Social Sciences): Math]. Ankara: Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Başkanlığı.

EARGED. (2009). *OBBS 2008 ilköğretim öğrencilerinin başarılarının belirlenmesi, Matematik, Fen ve Teknoloji, Sosyal Bilgiler, Türkçe, İngilizce raporu* [Determining the achievement level of primary school students in Math, Science and Technology, Social Sciences, Turkish and English]. Ankara: Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Başkanlığı.

EARGED. (2010). *PISA 2006 ulusal nihai raporu* [Final national report on PISA 2006 study]. Ankara: Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Başkanlığı.

Education Reform Initiative (ERI). (2010a). *Education monitoring report 2008*. Sabancı University, Istanbul: Istanbul Policy Center.

Education Reform Initiative (ERI). (2010b). *Education monitoring report 2009*. Sabancı University, Istanbul: Istanbul Policy Center.

Education Reform Initiative (ERI). (2011). Education monitoring report 2010. Sabancı University, Istanbul: Istanbul Policy Center.

Education Reform Initiative (ERI). (2012). *"4+4+4" düzenlemesi ile neler değişti? Yeni sisteme geçişte neler izlenmeleri?* [What has changed with the 4+4+4 regulation? What needs to be monitored with the new system?]. Sabancı University, Istanbul: Istanbul Policy Center.

Education Reform Initiative (ERI). (2013a). *Education monitoring report 2011*. Sabancı University, Istanbul: Istanbul Policy Center.

Education Reform Initiative (ERI). (2013b). *Education monitoring report 2012*. Sabancı University, Istanbul: Istanbul Policy Center.

Education Reform Initiative (ERI). (2014). *6 Şubat tarihinde TBMM'ye sunulan MEB yasa tasarısı üzerine on değerlendirmeler* [Preliminary analysis of the draft law on the Ministry of National Education presented to the Turkish Grand National Assembly]. Retrieved from http://erg.sabanciuniv.edu/node/1278

Eğinli, I. (2010). World Bank education policy: The impact of World Bank education projects (1997-2007) on Turkey (Doctoral and postdoctoral workshop on development and international organizations). Stockholm: World Bank's Annual Bank Conference on Development Economics.

Ergil, D. (2000). Identity crises and political instability in Turkey. *Journal of International Affairs-Colombia University*, *54* (1), 43-62.

Ergüder, U. (2013). The indispensable role of education for the centennial goals of the Turkish Republic. *Turkish Policy Quarterly*, 12 (2), 49-63.

Esme, I. (2002). *Avrupa Birliği bağlamında eğitimde yeni yönelimler ve özel okullar* [New directions in education and private schools in the context of European Union]. Antalya: Özel Okullar Derneği.

Estrella, M., Blauert, J., Campilan, D., Gaventa, J., Gonsalves, J., Guijt, I., Johnson, D.A., & Ricafort, R. (Eds.). (2000). *Learning from Change: Issues and experiences in participatory monitoring and evaluation*. Ottawa, Canada: International Development Research Centre.

European Commission. (2006a). Capacity building support for the Ministry of National Education (Project number TR 06 03 11). Retrieved from http://ec.europa.eu/enlargement/fiche_projet/document/TR%2006%2003%2011%20Capacity%20Building%20Support%20for%20the%20MoNE.pdf

European Commission. (2006b). Turkey 2005 progress report. Retrieved from http://ab.gov.tr/files/AB_Iliskileri/Tur_En_Realitons/Progress/Turkey_Progress_Report_2005.pdf

European Commission. (2007). Support to basic education program (SBEP): Discussion paper. Retrieved from http://tedp.meb.gov.tr/doc/Pubs/07.01.29%20White%20Paper%20Final%20Draft%20January%202007.pdf

European Commission. (2010). Public Financial Management and Control Law No. 5018: Terms of Reference. Retrieved from www.sipu.se/media/35615/tor.pdf

European Commission. (2012). Turkey 2011 progress report. Retrieved from http://ab.gov.tr/files/AB_Iliskileri/AdaylikSureci/IlerlemeRaporlari/tr_rapport_2011_en.pdf

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation. Alternative approaches and practical guidelines* (3rd ed.). New York: Pearson Education, Inc.

Fitzpatrick, J. L. (2012). An introduction to context and its role in evaluation practice. *New Directions for Evaluation*, *2012* (135), 7-24.

Fleischer, D. N., & Christie, C. A. (2009). Evaluation use results from a survey of US American Evaluation Association members. *American Journal of Evaluation*, *30*(2), 158-175.

Fowler, F. C. (2009). *Policy studies for educational leaders: An introduction*. Boston: Pearson.

Furubo, J. E., Rist, R., & Sandahl, R. (Eds.). (2002). *International atlas of evaluation*. New Brunswick, London: Transaction Publishers.

Garcia-Iriarte, E., Suarez-Balcazar, Y., Taylor-Ritzler, T., & Luna, M. (2011). A catalyst-for-change approach to evaluation capacity building. *American Journal of Evaluation*, 32(2), 168-182.

Gasper, D. (2000). Evaluating the logical framework approach: towards learning-oriented development evaluation. *Public Administration and Development,* 20, 17–28.

Gershberg, A. (2005). Towards an education decentralization strategy for Turkey. In World Bank (Eds.), *Turkey – Education sector study* (Report No. 32450-TU) (pp. 148-154).

Giovannini, E. (2009). Bringing statistics to citizens: A 'must' to build democracy in the XXI century. In M. Segone (Eds.), *Country-led monitoring and evaluation systems: Better evidence, better policies, better development results* (pp. 135-157). New York, NY: UNICEF.

Gök, F. (2011). Opening welcome address at the XIV World Congress of Comparative Education Societies, Istanbul, Turkey. In S. Majhanovich, C. Fox, & F. Gok, (Eds.), *Bordering, Re-Bordering and New Possibilities in Education and Society* (pp. 6-8). New York: Springer.

Government Action Plan, Republic of Turkey. (2003). 59th government action plan. Retrieved from http://www.tbmm.gov.tr/hukumetler/HP59.htm

Government Action Plan, Republic of Turkey. (2007). 60th government action plan. Retrieved from http://ekutup.dpt.gov.tr/plan/ep2008.pdf

Government Action Plan, Republic of Turkey. (2011). 61st government action plan. Retrieved from http://www.basbakanlik.gov.tr/Forms/pgGovProgramme.aspx

Greene, J. C. (2005). Context. In S. Mathison (Eds.), *Encyclopedia of evaluation* (pp. 82–84). Thousand Oaks, CA: Sage Publications Inc.

Greene, J. C. (2006). Evaluation, democracy, and social change. In I. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The Sage handbook of evaluation.* (pp. 118-140). Thousand Oaks, CA: Sage Publications Inc.

Grisay, A., & Griffin, P. (2006). What are the main cross-national studies? In K. N. Ross & I. J. Genevois (Eds.), *Cross-national studies of the quality of education: Planning their design and managing their impact* (pp. 67-103). Institute for Educational Planning: UNESCO.

Grubb, W. N., & Lazerson, M. (2004). *The education gospel: The economic power of schooling.* Cambridge: Harvard University Press.

Grudens-Schuck, N. (2003). The rigidity and comfort of habits: A cultural and philosophical analysis of the ups and downs of mainstreaming evaluation. *New Directions for Evaluation*, 99, 23-32.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation.* Thousand Oaks, CA: Sage Publications, Inc.

Guba, E.G., & Lincoln, Y.S. (1998). Competing paradigms in qualitative research. In N.K. Denzin and Y.S. Lincoln (Eds.), *The landscape of qualitative research: Theories and issues* (pp. 195-220). Thousand Oaks, CA: Sage Publications.

Guendel, L. (2012). Evaluation, public policies, and human rights. *New Directions for Evaluation*, 134, 29-37.

Gür, B. S., & Çelik, Z. (2009). *Türkiye'de milli eğitim sistemi: Yapısal sorunlar ve öneriler [National education system in Turkey: Structural problems and recommendations].* Ankara: Siyaset, Ekonomi ve Toplum Araştırmaları Vakfı.

Gür, B. S. (2011). Önsöz [Preface]. In B. S. Gur (Eds.), *2000'li yıllar: Türkiye'de eğitim* [The 2000s: Education in Turkey], (pp.7-14). Istanbul: Meydan

Hancock, D. R., & Algozzine, B. (2011). *Doing case study research: A practical guide for beginning researchers* (2nd Ed.). New York: Teachers College Press.

Harvey, D. (2005). *A brief history of neoliberalism.* Oxford: Oxford University Press.

Hay, K. (2010). Evaluation field building in South Asia: Reflections, anecdotes, and questions. *American Journal of Evaluation,* 31(2), 222-231.

Hay, K. (2011, June). Building the field of evaluation in South Asia: A framework and ideas. Community of Evaluators, Working Paper Series, 3.

Heck, R. (2008). *Studying educational and social policy: Theoretical concepts and research methods.* New Jersey: Lawrence Erlbaum Associates.

Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, *24*(3), 293-314.

Hood, S., & Hopson, R. K. (2008). Evaluation roots reconsidered: Asa Hilliard, a fallen hero in the "Nobody knows my name" project, and African educational excellence. *Review of Educational Research*, *78*(3), 410-426.

Hood, S., Hopson, R. K., & Frierson, H. T. (2005). *The role of culture and cultural context: A mandate for inclusion, the discovery of truth and understanding in evaluative theory and practice*. Charlotte, NC: Information Age Publishing.

Hopson, R. (2001). Global and local conversations on culture, diversity, and social justice in evaluation: Issues to consider in a 9/11 era. *American Journal of Evaluation*, 22(3), 375-80.

Hopson, R., Kirkhart, K. E., & Bledsoe, K. L. (2012). Decolonizing evaluation in a developing world: Implications and cautions for equity-focused evaluation. In M. Segone (Eds.), *Evaluation for equitable results* (pp. 59-82). New York, NY: UNICEF.

Horton, D. (1999). Building capacity in planning, monitoring and evaluation: Lessons from the field. *Knowledge, Technology & Policy*, *11*(4), 152-188.

Horton, D. (2003). *Evaluating capacity development: Experiences from research and development organizations around the world*. Ottawa, Canada: International Development Research Centre.

House, E. R., & Howe, K. R. (2003). Deliberative democratic evaluation. In T. Kellaghan, D. L. Stufflebeam, & L. Wingate (Eds.), *The international handbook of evaluation*, (pp. 79-100). Netherlands: Springer.

Hoy, W. K., & Miskel, C. G. (2008). *Educational administration: Theory, research, and practice.* Boston: McGraw-Hill Higher Education.

Ilgın, Y. (2010). *Results-based monitoring and evaluation systems: Project of Monitoring and Evaluation in Development Programs in Turkey* [PowerPoint Slides]. Retrieved from http://ived.kalkinma.gov.tr/DocObjects/Download/10964/MEDP_Project_Closing _Workshop_September_29-Yilmaz_ILGIN_Eng.ppt

İnal, K., & Akkaymak, G. (2012). *Neoliberal transformations of education in Turkey*. New York: Palgrave Macmillan.

İnal, K. (2012). The educational policies of the AKP: The collapse of the public education in Turkey. In K. İnal & G. Akkaymak (Eds.), *Neoliberal transformations of education in Turkey* (pp. 17-30). New York: Palgrave Macmillan.

Johnson, K., Greenseid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use a review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, *30*(3), 377-410.

Joint Committee on Standards for Educational Evaluation. (2011). *The program evaluation standards* (3rd Ed.). Thousand Oaks, CA: Sage Publications Inc.

Kawakami, A. J., Aton, K., Cram, F., Lai, M. K., & Porima, L. (2008). Improving the practice of evaluation through indigenous values and methods: Decolonizing evaluation practice – Returning the gaze from Hawai'i and Aotearoa. In N. L. Smith & P. R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 219-242). New York, NY: Guilford.

Keser-Aschenberger, F. (2012). *Dynamics of policy formation in Turkey and the U.S.: A comparative case study of two reform initiatives* (Unpublished doctoral dissertation). Middle East Technical University, Ankara.

Khan, M. A. (1998). Evaluation capacity building an overview of current status: Issues and options. *Evaluation*, *4*(3), 310-328.

Kincheloe, J. L., & McLaren, P. (2011). Rethinking critical theory and qualitative research. *Key Works in Critical Pedagogy*, 285-326.

King, J. A. (2002). Building the evaluation capacity of a school district. *New Directions for Evaluation*, 93, 63-80.

King, J. A., & Volkov, B. (2005). A framework for building evaluation capacity based on the experiences of three organizations. *CURA Reporter*. Retrieved from http://www.cura.umn.edu/sites/cura.advantagelabs.com/files/publications/35-3-King-Volkov.pdf

King, J. A. (2007). Developing evaluation capacity through process use. *New Directions for Evaluation,* 116, 45-59.

King, J. A. (2008). Bringing evaluative learning to life. *American Journal of Evaluation*, 29 (2), 151-155.

King, J. A. (2010). Response to Evaluation Field Building in South Asia: Reflections, Anecdotes, and Questions. *American Journal of Evaluation*, 31(2), 232-237.

King, J. A., & Stevahn, L. (2012). *Interactive evaluation practice: Mastering the interpersonal dynamics of Program Evaluation*. Thousand Oaks, CA: Sage Publications, Inc.

Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, *88*, 5-23.

Kirkhart, K. E. (2005). Through a cultural lens: Reflections on validity and theory in evaluation. *The Role of Culture and Cultural Context*, 21-38.

Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. *American Journal of Evaluation*, 31(3), 400-413.

Kocaman, O., & Güven, N. (2008). *Glossary of key terms in evaluation and results-based management*. Organisation for Economic Cooperation and Development. Retrieved from http://www.oecd.org/dac/evaluation/dcdndep/39059671.pdf

Kumar, A. K. (2010). A comment on "evaluation field building in South Asia: Reflections, anecdotes, and questions." *American Journal of Evaluation*, *31*(2), 238-240.

Kumar, R., & Hill, D. (2009). Introduction: Neoliberal capitalism and education. In D. Hill and R. Kumar (Eds.), *Global neoliberalism and education and its consequences,* (pp. 127-150). New York: Routledge.

Kushner, S., & Rotondo, E. (Eds.). (2012). *Evaluation voices from Latin America. New Directions for Evaluation*, 109. New Jersey: Wiley Periodicals Inc.

Küçükcan, T. (2003). State, Islam, and religious liberty in modern Turkey: Reconfiguration of religion in the public sphere. Retrieved from http://lawreview.byu.edu/archives/2003/2/KUC.pdf

LaFrance, J. (2004). Culturally competent evaluation in Indian country. *New Directions for Evaluation*, 102, 39-50.

LaFrance, J., & Nichols, R. (2008). Reframing evaluation: Defining an Indigenous evaluation framework. *Canadian Journal of Program Evaluation*, *23(*2), 13-32.

LaFrance, J., Nichols, R., & Kirkhart, K. E. (2012). Culture writes the script: On the centrality of context in indigenous evaluation. In D. J. Rog, J. L. Fitzpatrick, & R. F. Conner (Eds.), *Context: A framework for its influence on evaluation practice*. New Directions for Evaluation, 135, 59-74.

Lahey, R., & Nielsen, S. B. (2013). Rethinking the relationship among monitoring, evaluation and results-based management: Observations from Canada. *New Directions for Evaluation*, 137, 45-56.

Lee, Y. (2002). Public policy evaluation in Korea: In search for new direction. In J. E. Furubo, R. C. Rist, & R. Sandahl (Eds.), *International atlas of evaluation* (pp. 191- 205). Transaction Publishers.

Leviton, L. C. (2014). Some underexamined aspects of evaluation capacity building. *American Journal of Evaluation,* 35(1), 90-94.

Lincoln, Y. S., & Guba, E. G. (Eds.). (1985). *Naturalistic Inquiry*. Beverley Hills: Sage.

Lundgren, H., & Kennedy, M. (2009). Supporting partner country ownership and capacity in development evaluation. In M. Segone (Eds.), *Country-led monitoring and evaluation systems: Better evidence, better policies, better development results* (pp. 77-95). New York, NY: UNICEF.

Mackay, K. (2002). The World Bank's ECB experience. *New Directions for Evaluation*, 93, 81-100.

Mackay, K. (2009). Building monitoring and evaluation systems to build government performance. In M. Segone (Eds.), *Country-led monitoring and evaluation systems: Better evidence, better policies, better development results* (pp. 169-187). New York, NY: UNICEF.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco, CA: Jossey-Bass.

Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, *10* (1), 35-57.

Martin, S. (2005). Evaluation, inspection and the improvement agenda: Contrasting fortunes in an era of evidence-based policy making. *Evaluation*, 11(4), 496–504.

Martinic, S. (2012). Emergent evaluation and educational reforms in Latin America. *New Directions for Evaluation*, 134, 17-28.

Mathur, K. (2009). Policy research organizations in South Asia. Working Papers Series of the Centre for Law and Governance, Jawaharlal University. Retrieved from http://www.jnu.ac.in/CSLG/WorkingPapers.html

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage Publications, Inc.

Mayne, J., Divorski, S., & Lemaire, D. (1999). Locating evaluation: Anchoring evaluation in the executive or the legislature, or both or elsewhere? In R. Boyle & D. Lemaire (Eds.), *Building effective evaluation capacity: Lessons from Practice* (pp. 23-43). Transaction Publishers.

Menon, S. (2011). Statement by Mrs. Saraswathi Menon. In United Nations Development Programme. (Eds.), *National evaluation capacities. Proceedings from the international conference* (pp. 12-14). Washington, D.C.: Evaluation Office. Retrieved from http://web.undp.org/evaluation/documents/NEC/nec_proceedings_series_1.pdf

Merriam, S. B. (1998). Case studies as qualitative research. *Qualitative Research and Case Study Applications in Education*, 26-43.

Merryfield, M. M. (1985). The challenge of cross-cultural evaluation: Some views from the field. *New Directions for Program Evaluation*, 25, 3-17.

Mertens, D. M. (2007). Transformative paradigm mixed methods and social justice. *Journal of Mixed Methods Research*, *1*(3), 212-225.

Mertens, D. (2008). Stakeholder representation in culturally complex communities: Insights from the transformative paradigm. In N. L. Smith & P. R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 41-60). New York, NY: Guilford.

Mertens, D. M. (2009). *Transformative research and evaluation*. The Guilford Press.

Mertens, D. M., & Russon, C. (2000). A proposal for the International Organization for Cooperation in Evaluation. *American Journal of Evaluation*, *21*(2), 275-283.

Mertens, D. M., & Hopson, R. K. (2006). Advancing evaluation of STEM efforts through attention to diversity and culture. *New Directions for Evaluation*, 109, 35-51.

Ministry of Development, Republic of Turkey. (2001). Eight (8th) five year development plan (2001– 2006). Retrieved from http://www.mod.gov.tr/en/plans/Eight%20Five-Year%20Development%20Plan%202001-2005.pdf

Ministry of Development, Republic of Turkey. (2006). Ninth (9th) national development plan (2007 – 2013). Retrieved from http://www.maliye.gov.tr/Lists/TabMenuIcerik/Attachments/106/9developmentplan.pdf

Ministry of Development, Republic of Turkey. (2013). Tenth (10th) national development plan (2014-2019). Retrieved from http://www.resmigazete.gov.tr/eskiler/2013/07/20130706M1-1.htm

Ministry of European Affairs, Republic of Turkey. (2001). The national programme for the adoption of the Acquis. Retrieved from http://www.abgs.gov.tr/index.php?p=195&l=2

Ministry of European Affairs, Republic of Turkey. (2003). The national programme for the adoption of the Acquis. Retrieved from http://www.abgs.gov.tr/index.php?p=196&l=2

Ministry of European Affairs, Republic of Turkey. (2008). The national programme for the adoption of the Acquis. Retrieved from http://www.abgs.gov.tr/index.php?p=42260&l=2

Ministry of Finance, Republic of Turkey. (2003). Public financial management and control act No. 5018. Retrieved from http://www.sgb.gov.tr/Mevzuat/01.Kanun/Public%20Financial%20Management%20and%20Control%20Law%20No.%205018.pdf

Ministry of National Education, Republic of Turkey. (1999). 16th National Council decisions. Retrieved from http://ttkb.meb.gov.tr/meb_iys_dosyalar/2012_06/06021346_16_sura.pdf

Ministry of National Education, Republic of Turkey. (2002). National education at the beginning of 2002. Retrieved from http://www.meb.gov.tr/Stats/apk2002ing/apage29_48.htm

Ministry of National Education, Republic of Turkey. (2006). 17th National Council decisions. Retrieved from http://ttkb.meb.gov.tr/meb_iys_dosyalar/2012_06/06021327_17_sura.pdf

Ministry of National Education, Republic of Turkey. (2009). Milli Eğitim Bakanlığı 2010-2014 stratejik planı [Ministry of National Education 2010-2014 Strategic Plan]. Retrieved from http://abdigm.meb.gov.tr/dokumanlar/stratejikplan.pdf

Ministry of National Education, Republic of Turkey. (2010). 18th National Council decisions. Retrieved from http://www.meb.gov.tr/duyurular/duyurular2010/ttkb/18Sura_kararlari_tamami.pdf

Ministry of National Education, Republic of Turkey. (2011). Activity report. Retrieved from http://sgb.meb.gov.tr/www/meb-2011-yili-faaliyet-raporu/icerik/52

Ministry of National Education, Republic of Turkey. (2012). Milli Eğitim Bakanlığı faaliyet raporu 2011 [Ministry of National Education activity report 2011]. Retrieved from http://sgb.meb.gov.tr/meb_iys_dosyalar/2012_07/31015733_meb_faaliyet_raporu_2011.pdf

Ministry of National Education, Republic of Turkey. (2013). Milli Eğitim Bakanlığı faaliyet raporu 2012 [Ministry of National Education activity report 2012]. Retrieved from

http://sgb.meb.gov.tr/meb_iys_dosyalar/2013_07/31123903_meb_faaliyet_raporu
_2012.pdf

Murphy, T., & Sazak, O. (2012). *Turkey's civilian capacity in post-conflict construction*. Istanbul: Istanbul Policy Center.

Neirotti, N. (2012). Evaluation in Latin America: Paradigms and practices. In S. Kushner & E. Rotondo (Eds.), *Evaluation voices from Latin America*. New Directions for Evaluation, 134, 7–16.

Nevo, D. (1982). The international context for research on evaluation. *Evaluation News*, 3(4), 73–75.

Nevo, D. (1983). The conceptualization of educational evaluation: An analytical review of the literature. *Review of Educational Research*, *53*(1), 117-128.

Nielsen, S. B., Lemire, S., & Skov, M. (2011). Measuring evaluation capacity: Results and implications of a Danish study. *American Journal of Evaluation*, 32(2), 324-344.

Okcabol, R. (2012). Gericileşmenin ve piyasalaşmanın şifreleri 635, 651, 652, 653, 666, 4+4+4 ve….. [Codes of reactionary and marketization laws no. 635, 651, 652, 653, 666, 4+4+4+ and…]. *Eğitim Bilim Toplum*, 10 (39), 28-41.

Organisation for Economic Development and Cooperation (OECD). (2006). *The challenge of capacity development: Working towards good practice*. Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD). (2009). *Better aid. Aid effectiveness: A progress report on implementing the Paris Declaration*. Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD). (2010). *Evaluation in development agencies: Better aid*. Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD). (2013). Education policy outlook: Turkey. Retrieved from http://www.oecd.org/edu/EDUCATION%20POLICY%20OUTLOOK%20TURK EY_EN.pdf

Özdemir, S. M. (2009). Eğitimde program değerlendirme ve Türkiye'de eğitim programlarını değerlendirme çalışmalarının incelenmesi [Curriculum evaluation in education and an analysis of curriculum evaluations in Turkey]. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, *6*(2), 126-149.

Özdemir, B. S., & Beltekin, N. (2012). Türk eğitim sisteminin dönüşümünde uluslararası aktörler: IMF ve Dünya Bankası örnekleri [The role of international actors in transforming Turkey's education system: The cases of the IMF and the World Bank]. *YYU Eğitim Fakültesi Dergisi*, 1, 33-55.

Patton, M. Q. (1985). Cross-cultural nongeneralizations. *New Directions for Program Evaluation*, 25, 93-96.

Patton, M. Q. (1998). Discovering process use. *Evaluation*, *4*(2), 225-233.

Patton, M. Q. (2001, November). Reflections on mainstreaming evaluation. Comments presented during session 790 at the annual meeting of the American Evaluation Association, St. Louis.

Patton, M. Q. (2002). *Qualitative research & evaluation methods*. Thousand Oaks, CA: Sage Publications, Inc.

Patton, M. Q. (2008). *Utilization-focused evaluation (4th Ed.).* Thousand Oaks, CA: Sage Publications, Inc.

Patton, M. Q. (2010, November). National evaluation capacity development. Comments presented during session 276 at the annual meeting of the American Evaluation Association, San Antonio.

Patton, M. Q. (2012). *Essentials of utilization-focused evaluation*. Thousand Oaks, CA: Sage Publications, Inc.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Thousand Oaks, CA: Sage Publications, Inc.

Picciotto, R. (2003). International trends and development evaluation: The need for ideas. *American Journal of Evaluation,* 24(2), 227-234.

Picciotto, R. (2007). The new environment for development evaluation. *American Journal of Evaluation*, 28(4), 509-521.

Picciotto, R. (2009). Evaluating development: Is the country right unit? In M. Segone (Eds.), *Country-led monitoring and evaluation systems: Better evidence, better policies, better development results* (pp. 32-55). New York, NY: UNICEF.

Picciotto, R. (2011). The logic of evaluation professionalism. *Evaluation*, *17*(2), 165-180.

Pihama, L., Cram, F., & Walker, S. (2002). Creating methodological space: A literature review of Kaupapa Maori research. *Canadian Journal of Native Education*, *26*(1), 30-43.

Philips, D. (2004). Towards a theory of policy attraction in education. In G. Steiner-Khamsi (Eds.), *The global politics of educational borrowing and lending*, (pp. 54-67). New York: Teachers College Press.

Porter, S. (2013). Moving beyond teaching people to fish. The role of voluntary organizations for professional evaluation in sustainable learning through innovation and intermediation. In M. Segone, & J. Rugh (Eds.), *Evaluation and civil society: Stakeholders' perspectives on national evaluation capacity development* (pp. 52-70). New York, NY: UNICEF.

Preskill, H. (2008). Evaluation's second act: Spotlight on learning. *American Journal of Evaluation*, *29*(2), 127-138.

Preskill, H., & Torres, R. T. (2000). The learning dimension of evaluation use. *New Directions for Evaluation*, (88), 25-37.

Rai, K. (2001, November). Mainstreaming logic model: vision and contradictions. Paper presented at the annual meeting of the American Evaluation Association, St. Louis.

Riddell, A. (1999). Evaluations of educational reform programmes in developing countries: Whose life is it anyway? *International Journal of Educational Development*, *19* (6), 383-394.

Rinne, R. (2008). The growing supranational impacts of OECD and the EU on national educational policies, and the case of Finland. *Policy Futures in Education*, 6(6), 665-680.

Rist, R. C. (Eds.). (1990). *Program evaluation and the management of government: Patterns and prospects across eight nations*. New Brunswick, NJ: Transaction Publishers.

Rist, R. C., Boily, M. H., & Martin, F. (2011). *Influencing change: Building evaluation capacity to strengthen governance*. Washington, D.C.: World Bank Publications.

Rog, D. J. (2012). When background becomes foreground: Toward context-sensitive evaluation practice. *New Directions for Evaluation*, 135, 25-40.

Rogers, P. J. (2005). Accountability. *Encyclopedia of evaluation*, 2-4. Thousand Oaks, CA: Sage Publications, Inc.

Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (2003). *Evaluation: A systematic approach*. Thousand Oaks, CA: Sage Publications, Inc.

Russ-Eft, D. F., & Preskill, H. (2009). *Evaluation in organizations: A systematic approach to enhancing learning, performance and change* (2nd ed.). Basic Books.

Russon, C., & Russon, K. (Eds.). (2000). *The annotated bibliography of international program evaluation*. Dordrecht, The Netherlands: Kluwer.

Russon, C., & Russon, G. (2005). *International perspectives on evaluation standards* (No. 104). San Francisco: Jossey-Bass.

Sanders, J. R. (1994). *The Joint Committee on Standards for Educational Evaluation. The program evaluation standards: How to assess evaluations of educational programs*. Thousand Oaks, CA: Sage Publications, Inc.

Sanders, J. R. (2003). Mainstreaming evaluation. *New Directions for Evaluation*, 99, 3-6.

Şaşmaz, A. (2013). To which direction does the education policy of AK Party change? *ResearchTurkey*, 2(2), 40-47.

Scheurich, J. J., & Young, M. D. (1997). Coloring epistemologies: Are our research epistemologies racially biased? *Educational Researcher*, *26*(4), 4-16.

Schiavo-Campo, S. (2005). *Building country capacity for monitoring and evaluation in the public sector: Selected lessons of international experience*. Washington, D.C.: World Bank Publications.

Schwandt, T., Kosheleva, N., Segone, M., Naidoo, I. A., Carden, F. McKegg, K., Ofir, Z., & Miyoshi, K. (2013). International perspectives on evaluation. *American Journal of Evaluation*, 34(4), 567-589.

Scriven, M. S. (1967). *The methodology of evaluation* (Perspectives of curriculum evaluation, and AERA monograph series on curriculum evaluation No. 1). Chicago: Rand NcNally.

Scriven, M. (1991). *Evaluation thesaurus*. Thousand Oaks, CA: Sage Publications, Inc.

Scriven, M. S. (1994). Evaluation as a discipline. *Studies in Educational Evaluation*, 20(1), 147-166.

Segone, M., & Ocampo, A. (2006). *Creating and developing evaluation organizations: Lessons learned from Africa, Americas, Asia Australasia and Europe*. Lima, Peru: International Organisation for Cooperation in Evaluation (IOCE).

Segone, M. (Eds.). (2008). *Bridging the gap: The role of monitoring and evaluation in evidence-based policy making*. New York, NY: UNICEF.

Segone, M. (Eds.). (2009). *Country-led monitoring and evaluation systems: Better evidence, better policies, better development results*. New York, NY: UNICEF.

Segone, M., & Pron, N. (2008, May). The role of statistics in evidence-based policy making. *United Nations Statistical Commission and Economic for Europe Conference of European Statisticians*. New York, NY: UNICEF.

Segone, M., Sakvarelizde, G., & Vadnais, D. (2009). Strengthening country data collection systems: The role of the multiple indicator cluster surveys. In M. Segone (Eds.), *Country-led monitoring and evaluation systems: Better evidence, better policies, better development results* (pp. 238-251). New York, NY: UNICEF.

SenGupta, S., Hopson, R., & Thompson-Robinson, M. (2004). Cultural competence in evaluation: An overview. *New Directions for Evaluation*, 102, 5-19.

Çelik, Z., Boz, N., Gümüş, S., & Taştan, F. (2013). *4+4+4 eğitim reformunu izleme raporu* [Monitoring report on the 4+4+4 educational reform]. Ankara: Egitim Bir Sen.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin Company.

Simon, C. A. (2007). *Public policy: Preferences and outcomes*. New York: Pearson Longman.

Şişman, M., & Turan, S. (2003). Eğitimde yerelleşme ve demokratikleşme çabaları: Teorik bir çözümleme [Efforts to localize and democratize education: Theoretical analysis]. *Kuram ve Uygulamada Eğitim Yönetimi*, 34, 300-315.

Şişman, M. (2011). 2000'li yıllarda Türkiye eğitim sisteminin yönetimi [Administration of Turkish education system in the 2000s]. In B. S. Gur (Eds.), *2000'li yıllar: Türkiye'de eğitim* [The 2000s: Education in Turkey] (pp. 77-124). Istanbul: Meydan.

Smith, N., Brandon, P., Hwalek, M., Kistler, S., Labin, S., Rug, J., Thoma, V., & Yarnal, L. (2011). Looking ahead: The future of evaluation. *American Journal of Evaluation*, 32, 565-599.

Smith, L. T. (2012). *Decolonizing methodologies: Research and indigenous peoples (2nd Ed.)*. London: Zed Books.

Soudien, C. (2011). "Quality's 'others'?" The politics of bordering and re-bordering our educational standards. In S. Majhanovich, C. Fox, & F. Gök, (Eds.), *Bordering, Re-Bordering and New Possibilities in Education and Society* (pp. 15-30). New York: Springer.

Sridharan, S., & De Silva, S. (2010). On ambition, learning, and co-evolution: Building evaluation as a field. *American Journal of Evaluation*, 31(2), 246-251.

Stake, R. E. (1983). *Program evaluation, particularly responsive evaluation*. Netherlands: Springer.

Stake, R. E. (2005). Case studies. In N. K. Denzin, & Y. S. Lincoln (Eds.). *The SAGE handbook of qualitative research* (pp. 435-454). Thousand Oaks, CA: Sage Publications, Inc.

Stake, R. E. (2010). *Qualitative research: Studying how things work*. Guilford Press.

Steiner-Khamsi, G. (2003). The politics of league tables. *Journal of Social Science Education*, 2(1). Retrieved from http://www.jsse.org/jsse/index.php/jsse/article/view/470

Steiner-Khamsi, G. (2004). Globalization in education: Real or imagined? In G. Steiner-Khamsi (Eds.), *The global politics of educational borrowing and lending*, (pp. 54-67). New York: Teachers College Press.

Steiner-Khamsi, G. (2014, March). Standards are good for business: A critical examination of global reform packages. Group panel session presented at the 58th annual conference of the Comparative and International Education Society, Toronto, Canada.

Stockdill, S., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation*, 93, 7-26.

Stout, S. (2010). *State Planning Department Monitoring and Evaluation Development Project: Existing capacities and gaps in Turkey's M&E system* [PowerPoint slides]. Retrieved from ived.kalkinma.gov.tr/DocObjects/.../Existing_capacity_and_gaps.ppt

Strauss, A., & Corbin, J. (1994). Grounded theory methodology. In N. K. Denzin, & Y. S. Lincoln (Eds.). *The SAGE handbook of qualitative research* (pp. 273-285). Thousand Oaks, CA: Sage Publications, Inc.

Stufflebeam, D.L. (1968). *Evaluation as enlightenment for decision-making*. Columbus, Ohio: Evaluation Center.

Stufflebeam, D.L. (2002). Institutionalizing evaluation checklist. Western Michigan University. Retrieved from http://www.wmich.edu/evalctr/archive_checklists/institutionalizingeval.pdf

Stufflebeam, D. L. (2004). The 21st-century CIPP model. *Evaluation roots*, 245-266.

Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications* (Vol. 3). San Francisco: Jossey-Bass.

Tarsilla, M. (2012). *From building evaluation capacity to sustaining evaluation capacity development: The cases of Democratic Republic of Congo, Niger, and South Africa* (Unpublished doctoral dissertation). Western Michigan University, Michigan.

Taylor-Powell, E., & Boyd, H. H. (2008). Evaluation capacity building in complex organizations. *New Directions for Evaluation*, 120, 55-69.

Terzi, C. (2005). *Uyum surecinde Türkiye eğitim politikalarının Avrupa Birliği eğitim politikaları doğrultusunda değerlendirilmesi* [An analysis of Turkey's educational policies within the context of European Union educational policies in the process of harmonization] (Unpublished doctoral dissertation). Anadolu University, Eskisehir.

TEPAV. (2008). PISA sonuçları ışığında Türkiye'nin rekabet gücünün değerlendirilmesi [An analysis of Turkey's competitive power in light of PISA results]. Retrieved from http://www.tepav.org.tr/upload/files/1271251457r2163.PISA_Sonuclari_Isiginda_Turkiye___nin_Rekabet_Gucunun_Degerlendirilmesi.pdf

TEPAV. (2013). Onuncu kalkınma planı hakkında bir değerlendirme [An analysis of the tenth national development plan]. Retrieved from http://www.tepav.org.tr/upload/files/1376393421-7.Onuncu_Kalkinma_Plani_Hakkinda_Bir_Degerlendirme.pdf

Theodoulou, S. Z., & Kofinis, C. (2004). *The art of the game: Understanding American public policy making*. California: Thomson-Wadsworth.

Torres, R. T., & Preskill, H. (2001). Evaluation and organizational learning: Past, present, and future. *American Journal of Evaluation*, *22*(3), 387-95.

Turan, S. (2000). John Dewey's report of 1924 and his recommendations on the Turkish educational system revisited. *History of Education*, 29(6), 543-555.

Turan, S. (2008). Eğitim felsefesi ve çağdaş eğitim sistemleri [Educational philosophy and modern education systems]. In A. Boyacı (Eds.), *Eğitim sosyolojisi ve felsefesi* [Educational sociology and philosophy] (pp. 1-21). Eskişehir: Anadolu Üniversitesi Acık Öğretim Fakültesi Yayını, 973.

Türk, E., Yalçın, M., & Ünsal, N. (2006). *Milli Eğitim Bakanlığı yöneticilerinin görüşlerine dayalı stratejik planlama araştırması* [A research on the Ministry of National Education officials' perceptions of strategic planning]. Ankara: Milli Eğitim Bakanlığı Strateji Geliştirme Başkanlığı.

United Nations Children's Fund (UNICEF). (2010). Evaluation for children: Developing national capacities for country-led evaluation systems. Retrieved from http://www.unicef.org/evaluation/files/DevelopingCapacities_CountryEvaluation Systems_final.pdf

United Nations Development Programme (UNDP). (2010). *Assessment of development results: Turkey. Evaluation of UNDP contribution to development results in Turkey*. Washington, D.C.: Evaluation Office.

United Nations Development Programme (UNDP). (2011). *Proceedings from the international conference 2011.National evaluation capacities.* Washington, D.C.: Evaluation Office. Retrieved from http://web.undp.org/evaluation/documents/NEC/nec_proceedings_series_1.pdf

United Nations Development Group (UNDG). (2011). Results-based management handbook. Retrieved from http://www.undg.org/docs/12316/UNDG-RBM%20Handbook-2012.pdf

Urgent Action Plan, Republic of Turkey. (2003). 58. Hükümet acil eylem planı [the 58th government urgent action plan]. Retrieved from http://www.linux.org.tr/wp-content/uploads/2010/04/AcilEylemPlani.pdf

Valverde, G. A. (2014, March). Educational quality: Global politics, comparative inquiry and opportunities to learn. Presidential address at the 58th annual conference of the Comparative and International Education Society, Toronto, Canada.

Vidueira, P., Diaz-Puente, J. M., & Afonso, A. (2013). The worldwide expansion of evaluation: A world of possibilities for rural development. *Cuadernos de Desarrollo Rural*, 10(70), 159-180.

Vision 2023 for Turkey: Priorities for skills. (2013, October). [PowerPoint Slides]. Retrieved from http://www.etf.europa.eu/eventsmgmt.nsf/(getAttachment)/C5138B787A1E9C64 C1257C0D00253DB7/$File/Turkey%20Vision%202023.pptx

Wadsworth, Y. (2001). Becoming responsive and some consequences for evaluation as dialogue across distance. In J. C. Greene and T. A. Abma (Eds.), *Responsive Evaluation. New Directions for Evaluation*, 92. San Francisco: Jossey-Bass.

Weiss, C. H. (1972). Utilization of evaluation: Toward comparative study. *Evaluating action programs: Readings in social action and education*, 318-326.

Weiss, C. H. (1973). Between the cup and the lip. *Evaluation*, *1*(2), 49-55.

Weiss, C. H. (Eds.). (1977). *Using social research in public policy making.* Lexington, MA: Lexington Books.

Weiss, C. H. (1979). The many meanings of research utilization. *Public administration review*, *39*(5), 426-431.

Weiss, C. H., & Bucuvalas, M. J. (1980). *Social science research and decision-making*. New York: Columbia University Press.

Weiss, C. H. (1987). Evaluating social programs: What have we learned?. *Society*, *25*(1), 40-45.

Weiss, C. H. (1998a). *Evaluation* (2nd Ed). New Jersey: Prentice Hall.

Weiss, C. H. (1998b). Have we learned anything new about the use of evaluation?. *American Journal of Evaluation*, *19*(1), 21-33.

Weiss, C. H., Murphy-Graham, E., & Birkeland, S. (2005). An alternate route to policy influence how evaluations affect DARE. *American Journal of Evaluation*, *26*(1), 12-30.

Williams, D. D., & Hawkes, M. L. (2003). Issues and practices related to mainstreaming evaluation: Where do we flow from here?. *New Directions for Evaluation*, 99, 63-83.

World Bank. (2002). *2002 annual report on evaluation capacity development*. Washington, D.C.: World Bank Publications. Retrieved from http://siteresources.worldbank.org/EXTEVACAPDEV/Resources/4585672-1252437578683/AREC_2002.pdf

World Bank. (2004). *Evaluation capacity development: OED self-evaluation*. Washington, D.C.: World Bank Publications. Retrieved from http://siteresources.worldbank.org/INTEVACAPDEV/Resources/4585664-1253892707805/ecd_self_evaluation.pdf

World Bank. (2005). *Turkey: Education sector study*. (Report No 32450-TU). Washington, D.C.: The World Bank.

World Bank. (2012). Implementation completion and results report on a loan in the amount of Euro 80 million (US$ 2014 million equivalent) to the Republic of Turkey for a Secondary Education Project (Report No. ICR2264). Retrieved from http://documents.worldbank.org/curated/en/2012/12/17171935/turkey-secondary-education-project

World Bank. (2013a). *Country partnership strategy for the Republic of Turkey for the period 2012-2015*. Retrieved from http://documents.worldbank.org/curated/en/2013/01/17369150/turkey-country-partnership-strategy-period-2012-2015

World Bank. (2013b). *Promoting excellence in Turkey's schools* (Report No. 77722-TR). Retrieved from http://documents.worldbank.org/curated/en/2013/03/18023851/promoting-excellence-turkeys-schools

Yenice, E. (2006). Kamu kesiminde performans değerlendirmesi [Performance evaluation in the public sector]. *Maliye Dergisi*, 150, 122-132.

Yin, R. K. (1994). *Case study research: Design and methods* (2nd Ed.). Thousand Oaks, CA: Sage.

Young, J. (2005). Research, policy and practice: why developing countries are different. *Journal of International Development*, *17*(6), 727-734.

Yüksel, I., & Sağlam, M. (2012). *Eğitimde program değerlendirme* [Curriculum evaluation in education]. Ankara: Pegem Press.

Appendix A. Interview Protocols

Semi-Structured Interview Protocol for Government Officials

     In English. Good afternoon. Thank you for agreeing to participate in this interview. My name is Hanife Cakici, a Ph.D. candidate from the University of Minnesota in the United States. This interview is a part of my doctoral research study that aims to explore educational decision-making practices, and determinants of effective educational programs and policies. First, let me explain that I will take notes during the interview. Second, may I record our interview with an audio recorder? You are free to say "no" now or at any point in our conversation. The purpose of an audio recording is only to aid in note-taking; any recording made will not be included with the report or shared with anyone. Is this acceptable? Please feel free to share your point of view honestly. Both negative and positive comments are welcomed; sometimes the negative comments are the most helpful. Do you have any questions for me before we start?

Part One: Educational decision making practices

1. Can you briefly describe your job/position within the Ministry?
   a. How does your role—directly or indirectly—relate to educational programs and policies?
2. How would you describe the culture of this Ministry?
3. Can you explain how decisions are made regarding educational programs and policies at the Ministry?
   a. How are tasks typically structured?
   b. Who holds power in educational decision making process?
4. What factors (values, politics, personalities etc.) affect educational decision making?
   a. What values, cultural issues, traditions, personalities, expectations in the educational decision making context are crucial to understand how policy and programmatic decisions are made and executed?
   b. Are there any topics that are simply off-limits for discussion in this Ministry?
5. What are some of the most current educational programs and policies?
   a. How long have they been in operation?
   b. What do you think these programs or policies intend to achieve?
   c. How do you know that they are achieving their intended outcomes?
   d. To what extent are these programs' and policies' goals and objectives specified?

Part Two: Experience with evaluation

1. What data is currently collected on educational programs and policies?
   a. How is it used?
   b. To what extent do people here routinely use data as part of ongoing educational activities?
2. When you think about program evaluation, what word comes to your mind?

3. What previous evaluation experiences have you had?
   a. What are the examples of evaluations that you consider useful? What made them useful?
4. To what extent has Ministry staff received training in program evaluation, whether formal or informal?
   a. Has the Ministry engaged in evaluation capacity building or continuous improvement processes?
   b. How would you characterize staff's experiences?
5. How would you say program evaluation is viewed here?
   a. Are there any evaluation champions in the Ministry?
   b. To what extent is there potential resistance to evaluation?

Part Three: Evaluation's perceived influence on decisions

1. What role do you see for evaluation in educational decision making context?
   a. How much influence do you expect evaluation to have on educational decisions?
   b. How much do (would) evaluation results affect the continued work of the educational programs and policies?
   c. What ethical, cultural and political issues require attention to understand the role of evaluation in educational context?
2. What needs to be done to achieve the desired level of influence?
   a. What structure exists to use the results?
3. What kinds of barriers exist for utilizing evaluations to support educational decision-making?
4. What kinds of opportunities exist for utilizing evaluations to support educational decision-making?

In Turkish. Görüşme isteğimi geri çevirmediğiniz için çok teşekkür ederim. Halen Amerika Birleşik Devletleri'nde Minnesota Üniversitesi'nde devam etmekte olan doktora eğitimim kapsamında yürüttüğüm tez çalışması için bu görüşmeyi yapmak istedim sizinle. "İzleme/Değerlendirme Çalışmalarının Türkiye'de Eğitim Kararları için Değer Önermesi" başlıklı bu doktora tezi ülkemizde eğitimde karar alma mekanizmalarında izleme/değerlendirme çalışmalarının rolünü ve önemini araştırmayı amaçlamaktadır. Bulunduğunuz konum itibariyle, tecrübe ve bilgi birikiminizin bu çalışmaya büyük bir katkı sağlayacağını düşünmekteyim. Bu çalışma Milli Eğitim Bakanlığı ve Kalkınma Bakanlığı'nda çalışan üst düzey bürokratlar ile eğitim fakültelerinde araştırma yapan deneyimli akademisyenleri içermektedir. Araştırmanın size ve bulunduğunuz kuruma hiçbir risk getirmesi beklenmemektedir. Bu bilimsel araştırma kapsamında kişisel bilgilerinizi gizli tutulacaktır. Daha fazla bilgi için ekte Katılımcı Bilgi ve Onay formunu bulabilirsiniz. Bu araştırma Minnesota Üniverste'sinin Etik kurulu tarafından onaylanmıştır. Başlamadan önce sormak istediğiniz bir soru var mı?

<u>Birinci Bölüm</u>: Eğitimsel faaliyetler ve uygulamalar

1. Kısaca Milli Eğitim Bakanlığı'ndaki görevinizden bahseder misiniz? Ne tür sorumluluklarınız var?
    a. Bakanlıktaki görevinizin (direk ya da dolaylı olarak) eğitim politikaları ve projeleri ile bağlantısı nedir?
2. Eğitim sistemini etkileyen kararların (eğitim politikaları ve projeleri ile bakanlığın yapısal işleyişini etkileyen hükümlerin) normal şartlarda nasıl alındığını anlatabilir misiniz?
    a. Kararları kim nasıl biçimlendiriyor?
    b. Bu süreçte bakanlık birimlerine ne gibi rol düşüyor?
3. Eğitimde karar alma sürecini etkileyen faktörler sizce nelerdir?
    a. Hangi değer yargıları, kültürel, ekonomik ve sosyal olgular, gelenekler, yaşam tarzları, toplumsal ya da bireysel beklentiler eğitim politikaları ve projelerinin yapılanmasına ve uygulanmasına etki ediyor? Örnek verebilir misiniz?
    b. Tecrübenize dayanarak, şu anki mevcut düzende, eğitim kararları kapsamında konuşulması, karara bağlanması ya da uygulanması mümkün olmayacak, yasak-bölge olarak adlandıracağınız konu başlıkları var mı?
4. Bulunduğunuz birimi yakından ilgilendiren güncel eğitim politikaları ve projelerinden bazıları nelerdir?
    a. Ne zamandır bu politika/projeler yürürlükte?
    b. Bu politika ya da projelerin (herhangi biri ya da hepsi) hedefleri nelerdir?
    c. Bu politika/projelerin hedefleri ne derece belirlenmiş, tanımlanmış ve şekillenmiştir?
    d. Bu politika ve projelerin hedeflerine ulaşıp ulaşmadığını biriminiz (ya da bakanlık) nasıl anlıyor?

<u>İkinci Bölüm</u>: İzleme / Değerlendirme Alanı Hakkında Tecrübe ve Birikimler

6. Genelde eğitim politikaları ve projeleri ile ilgili ne tür araştırmalar yapılıyor?

      a. Ne tür veriler toplanıyor?

      b. Bu veriler/araştırma sonuçları nasıl kullanılıyor?

      c. Bakanlık çalışanları bu verilere/araştırmalara politika uygulama ve geliştirme sürecinde ne sıklıkla başvuruyorlar?

7. Politika ya da proje izleme/değerlendirmesi deyince aklınıza ne geliyor (bir kelime, bir resim ya da figür olabilir). Lütfen düşündüğünüz kelime ya da figürü anlatır mısınız?

8. İzleme/değerlendirme çalışmaları ile ilgili geçmiş deneyim ve kazanımlarınız nelerdir?

      a. Tecrübelerinize dayanarak ne tür değerlendirme çalışmalarını faydalı buldunuz şimdiye kadar? Bunları faydalı yapan faktörler nelerdir?

9. Bakanlık personeliyle birlikte ya da siz bireysel olarak politika/proje izleme/değerlendirme eğitimi (değerlendirme kapasitesi geliştirme ya da kalite iyileştirme süreci vb.) aldınız mı?

      a. Eğitimi kim ne zaman verdi?

      b. Çalışma arkadaşlarınızın (ya da sizin) bu eğitime olan tepkisi nasıldı?

10. Bakanlıkta, politika/proje izleme/değerlendirmesi genel olarak nasıl algılanıyor (görülüyor)?

      a. Bakanlık bünyesinde politika/proje izleme/değerlendirme konusunda öncü olan kişi ya da birimler var mi? Hangileri?

<u>Üçüncü Bölüm</u>: İzleme/değerlendirmenin karar alma sürecinde rolü

11. Şu anki mevcut durumda, politika/proje izleme/değerlendirme çalışmaları, Türkiye'de eğitimde karar alma sürecinde ne tarz bir rol oynuyor?

      a. Bu konuda geleceğe dönük beklentileriniz nelerdir?

      b. Türkiye'de, değerlendirmenin eğitim alanında rolünü anlamak için ne tarz etik, kültürel, siyasi ya da ekonomik faktörleri göz önünde bulundurmamız gerekiyor?

12. İzleme/değerlendirme çalışmalarının eğitim kararlarında etkili olması için neye ihtiyaç var (neler yapılması gerekiyor)?

      c. Bakanlık bünyesinde bu etkiyi yaratabilecek ne tür sistemler ya da yapılar mevcut?

13. Eğitimde karar alma sürecinde izleme/değerlendirme sistemlerini ve uygulamalarını kullanmada ne tür engeller var?

14. Eğitimde karar alma sürecinde izleme/değerlendirme sistemlerini ve uygulamalarını kullanmada ne tür fırsatlar mevcut?

Semi-Structured Interview Protocol for Nongovernmental Respondents

In English. Good afternoon. Thank you for agreeing to participate in this interview. My name is Hanife Cakici, a Ph.D. candidate from the University of Minnesota in the United States. This interview is a part of my doctoral research study that aims to explore educational decision-making practices, and determinants of effective educational programs and policies. First, let me explain that I will take notes during the interview. Second, may I record our interview with an audio recorder? You are free to say "no" now or at any point in our conversation. The purpose of an audio recording is only to aid in note-taking; any recording made will not be included with the report or shared with anyone. Is this acceptable?

Please feel free to share your point of view honestly. Both negative and positive comments are welcomed; sometimes the negative comments are the most helpful.

Do you have any questions for me before we start?

Part One: Educational decision making practices

1. Can you briefly describe your research interests (or mission of your organization)?
   a. How do your research interests—directly or indirectly—relate to educational programs and policies?
2. To what extent are you involved with educational decision making?
   a. Have you worked with a governmental agency before?
      i. If yes, what did you do?
3. What are some of the most current educational programs and policies that you are aware of or conduct research or evaluation on?
   a. What do you think these programs or policies intend to achieve?
   b. How do you think they are achieving their intended outcomes?
4. In your opinion, how decisions are made regarding educational programs and policies in Turkey?
   a. Who holds power in educational decision making process?
5. What factors (values, politics, personalities etc.) affect educational decision making?
   a. What values, cultural issues, traditions, personalities, expectations in the educational decision making context are crucial to understand how policy and programmatic decisions are made and executed?
   b. Are there any topics that are simply off-limits for discussion in the educational context?

Part Two: Experience with evaluation

1. What data is currently collected on educational programs and policies?
   a. To what extent do educational officials routinely use data as part of ongoing educational activities?
2. When you think about evaluation, what word comes to your mind?
3. What previous evaluation experiences have you had?

       a. What are the examples of evaluations that you consider useful for educational policies and programs? What made them useful?

4. To what extent have the educational officials engaged in evaluation capacity building or continuous improvement processes?

5. How would you say evaluation is viewed at the educational decision making setting?
       a. Are there any evaluation champions in educational decision making setting that you are aware of?
       b. To what extent is there potential resistance to evaluation in educational decision making?

Part Three: Evaluation's perceived influence on decisions

1. What role do you see for evaluation in educational decision making context?
       a. How much influence do you expect the evaluation to have on educational decisions?
       b. How much do (would) evaluation results affect the continued work of the educational programs and policies?
       c. What ethical, cultural and political issues require attention to understand the role of program evaluation in educational context?

2. What needs to be done to achieve the desired level of influence?
       a. What structure exists to use the results?

3. What kinds of barriers exist for utilizing evaluations to support educational decision-making?

4. What kinds of opportunities exist for utilizing evaluations to support educational decision-making?

In Turkish. Görüşme isteğimi geri çevirmediğiniz için çok teşekkür ederim. Halen Amerika Birleşik Devletleri'nde Minnesota Üniversitesi'nde devam etmekte olan doktora eğitimim kapsamında yürüttüğüm tez çalışması için bu görüşmeyi yapmak istedim sizinle. "İzleme/Değerlendirme Çalışmalarının Türkiye'de Eğitim Kararları için Değer Önermesi" başlıklı bu doktora tezi ülkemizde eğitimde karar alma mekanizmalarında izleme/değerlendirme çalışmalarının rolünü ve önemini araştırmayı amaçlamaktadır. Bulunduğunuz konum itibariyle, tecrübe ve bilgi birikiminizin bu çalışmaya büyük bir katkı sağlayacağını düşünmekteyim. Bu çalışma Milli Eğitim Bakanlığı ve Kalkınma Bakanlığı'nda çalışan üst düzey bürokratlar ile eğitim fakültelerinde araştırma yapan deneyimli akademisyenleri içermektedir. Araştırmanın size ve bulunduğunuz kuruma hiçbir risk getirmesi beklenmemektedir. Bu bilimsel araştırma kapsamında kişisel bilgilerinizi gizli tutulacaktır. Daha fazla bilgi için ekte Katılımcı Bilgi ve Onay formunu bulabilirsiniz. Bu araştırma Minnesota Üniversite'sinin Etik kurulu tarafından onaylanmıştır. Başlamadan önce sormak istediğiniz bir soru var mı?

Birinci Bölüm: Eğitimsel faaliyetler ve uygulamalar

1. Kısaca araştırma yaptığınız alanlardan/konulardan bahseder misiniz?
   a. Araştırma yaptığınız alanların – direk ya da dolaylı olarak – eğitim politikaları ve projeleri ile bağlantısı nedir?
2. Ne derece eğitimde karar alma surecine dahil oldunuz bundan önce?
   a. Hükûmetle ya da bakanlıkla önceden çalıştınız mi (araştırma ya da araştırma dışı bağlamda)?
      i. Evet ize, ne tarz bir çalışma yaptınız?
3. Üzerinde şu an araştırma yaptığınız ya da araştırma alanlarınızı yakından ilgilendiren son dönemlerdeki bazı eğitim politika ya da projeleri nelerdir?
   a. Sizce, bu politika ya da projelerin (herhangi biri ya da hepsi) hedefleri nelerdir?
   b. Bu politika/projelerin hedefleri ne derece belirlenmiş, tanımlanmış ve şekillenmiştir?
   c. Bu politika ve projelerin hedeflerine ulaşıp ulaşmadığını sizce ilgili kişiler nasıl anlıyor?
4. Sizce, eğitim sistemini etkileyen kararlar (eğitim politikaları ve projeleri ile bakanlığın yapısal işleyişini etkileyen hükümlerin) normal şartlarda nasıl alınıyor?
   a. Kararları kim nasıl biçimlendiriyor?
5. Eğitimde karar alma sürecini etkileyen faktörler sizce nelerdir?
   a. Hangi değer yargıları, kültürel, ekonomik ve sosyal olgular, gelenekler, yaşam tarzları, toplumsal ya da bireysel beklentiler eğitim politikaları ve projelerinin yapılanmasına ve uygulanmasına etki ediyor? Örnek verebilir misiniz?

İkinci Bölüm: İzleme / Değerlendirme Alanı Hakkında Tecrübe ve Birikimler

1. Genelde eğitim politikaları ve projeleri ile ilgili ne tür araştırmalar yapılıyor?
   a. Ne tür veriler toplanıyor?
   b. Bu veriler/araştırma sonuçları nasıl kullanılıyor?

  c. Bakanlık çalışanları bu verilere/araştırmalara politika uygulama ve geliştirme sürecinde ne sıklıkla başvuruyorlar?

2. Politika ya da proje izleme/değerlendirmesi deyince aklınıza ne geliyor (bir kelime, bir resim ya da figür olabilir). Lütfen düşündüğünüz kelime ya da figürü anlatır mısınız?
3. İzleme/değerlendirme çalışmaları ile ilgili geçmiş deneyim ve kazanımlarınız nelerdir?
  a. Tecrübelerinize dayanarak ne tür değerlendirme çalışmalarını faydalı buldunuz şimdiye kadar? Bunları faydalı yapan faktörler nelerdir?
4. Bakanlık personelinin politika/proje izleme/değerlendirme eğitimi (değerlendirme kapasitesi geliştirme ya da kalite iyileştirme süreci vb.) konusunda ne düşünüyorsunuz?
5. Türkiye'de eğitimde karar alma surecinde, politika/proje izleme/değerlendirmesi genel olarak nasıl algılanıyor (görülüyor)?
  a. Bakanlık bünyesinde politika/proje izleme/değerlendirme konusunda öncü olan kişi ya da birimler var mi? Hangileri?

<u>Üçüncü Bölüm</u>: İzleme/değerlendirmenin karar alma sürecinde rolü

1. Şu anki mevcut durumda, politika/proje izleme/değerlendirme çalışmaları, Türkiye'de eğitimde karar alma sürecinde ne tarz bir rol oynuyor?
  a. Bu konuda geleceğe dönük beklentileriniz nelerdir?
  b. Türkiye'de, değerlendirmenin eğitim alanında rolünü anlamak için ne tarz etik, kültürel, siyasi ya da ekonomik faktörleri göz önünde bulundurmamız gerekiyor?
2. İzleme/değerlendirme çalışmalarının eğitim kararlarında etkili olması için neye ihtiyaç var (neler yapılması gerekiyor)?
  a. Bakanlık bünyesinde bu etkiyi yaratabilecek ne tür sistemler ya da yapılar mevcut?
3. Eğitimde karar alma sürecinde izleme/değerlendirme sistemlerini ve uygulamalarını kullanmada ne tür engeller var?
4. Eğitimde karar alma sürecinde izleme/değerlendirme sistemlerini ve uygulamalarını kullanmada ne tür fırsatlar mevcut?

In English

You are invited to be in a research study of exploring the role and utility of evaluation in educational decision domain in Turkey. You were selected as a possible participant because of the formal position you occupy and the knowledge you have related to Turkish educational decision making, research, and evaluation. I ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by: Hanife Cakici, Ph.D. Candidate, Department of Organizational Leadership, Policy, and Development, University of Minnesota.

Background Information

The purpose of this study is to explore the value of evaluation for educational decision making in Turkey. Specifically, the proposed research seeks to understand (1) the extent to which evaluation is being used in educational decision domain, (2) current conceptualization of evaluation within Turkish context, (3) the need and capacity to conduct and use evaluations, (4) the influence of contextual factors on evaluation utilization for educational programs and policies.

Procedures

If you agree to be in this study, you will be asked to participate in a face-to-face in-depth interview once that will last for about 45 minutes to an hour. I will audiotape the interview and later type up the conversation. This way I will be better able to remember what we discussed and the exact words you used during the interview.
Risks and Benefits of being in the Study

There are no risks associated with participating in the interview. A benefit to participate in the interview will be to share your experiences, opinions, and perspectives related to the utility of evaluation for decision making, and potentially contribute to the effectiveness of educational policies and programs that impact the lives of Turkish people.

Compensation

You will not receive any compensation for participating in the interview.

Confidentiality

The records of this study will be kept private. In any sort of report I might publish, I will not include any information that will make it possible to identify you. Research records will be stored securely and only I will have access to the written and audio records. Study data will be encrypted according to current University policy for protection of confidentiality.

Voluntary Nature of the Study

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with your organization. If you decide to participate, you are free to not answer any question or withdraw at any time without affecting those relationships.

Contacts and Questions

The researcher conducting this study is: Hanife Cakici. You may ask any questions you have now. If you have questions later, **you are encouraged** to contact her at, cakic002@umn.edu or at (90) 553 460-9613. You can also contact the student's adviser Jean A. King at 430F Wulling Hall, 86 Pleasant St. SE, Minneapolis, Minnesota 55455 USA, 001 (612) 626-1614.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota, 55455 USA, 001 (612) 625-1650.

You will be given a copy of this information to keep for your records.

If you sign this paper, you indicate that you understand what this research is about and you agree to participate in the interviews.

Statement of Consent

I have read the above information. I have asked questions and have received answers. I consent to participate in the study.

Signature: _____Date: _____

Signature of Investigator: _____Date: _____

In Turkish

Amerika Birleşik Devletleri'nde Minnesota Üniversitesi'nde eğitim gören doktor adayı Hanife Çakıcı "İzleme / Değerlendirme Çalışmalarının Türkiye'de Eğitim Kararları için Değer Önermesi " üzerine bir doktora araştırması yürütmektedir. Bulunduğunuz konum dolayısıyla ve sahip olduğunuz bilgi birikimi sebebiyle sizi bu araştırmaya katkıda bulunmak için davet ediyoruz. Kararınızdan önce araştırma hakkında sizi bilgilendirmek istiyoruz. Bu bilgileri okuduktan sonra araştırmaya katılmak isterseniz lütfen bu formu imzalayıp kapalı bir zarf içinde bize ulaştırınız.

Araştırmanın Arka Planı

Bu araştırmanın amacı Türkiye'de eğitimde karar alma sürecinde izleme/değerlendirme çalışmalarının rolünü ve yararını, anahtar paydaşların bakış açısından tespit etmektir. Özellikle, bu araştırma (1) Türkiye'de eğitimde karar alma sürecinde izleme/değerlendirme çalışmalarının ne derece ve hangi şekilde kullanıldığını, (2) izleme/ değerlendirmenin bu bağlamda nasıl kavramsallaştırıldığını, (3) izleme/değerlendirme çalışmalarına duyulan ihtiyaç ve bu çalışmaları yürütebilecek kapasite, (4) izleme /değerlendirme çalışmalarının uygulanması ve kullanılmasına ilişkin bağlamsal etkenlerin anlaşılmasına ve keşfedilmesine olanak sağlamayı amaçlamaktadır.

Prosedürler

Bu araştırmaya katılmayı kabul ettiğiniz takdirde sizden yaklaşık 45 dakika ile 1 saat sürecek olana yüz yüze bir görüşmeye katılmanız istenmektedir. Bu görüşme bir ses kayıt cihazı ile kayıt altına alınacaktır. Bu yolla görüşme sırasında konuşulanlar eksiksiz kağıda dökülebilecek ve gerekirse yeniden gözden geçirilebilecektir.

Araştırmanın Risk ve Yararları

Bu araştırmanın size risk getirmesi beklenmemektedir. Araştırmaya katılmanın size direk bir yararı olacağı saptanmasa da, paylaşacağınız bilgi ve birikiminiz, Türkiye'de eğitim politika ve projelerinin daha da verimli hale gelmesine katkıda bulunacaktır.

Araştırmaya katılma bedeli

Sizden ücret talep edilmemektedir ve size herhangi bir ödeme yapılmayacaktır.

Gizlilik

Bu araştırma bilimsel bir amaçla yapılmaktadır ve katılımcı bilgilerinin gizliliği esas tutulmaktadır. Ses kayıtlarında katılımcının ismi yerine bir numara kullanılacaktır. Ses kasetleri araştırma süresince kilitli bir dolapta muhafaza edilip araştırma sona erdiğinde silineceklerdir. Yayınlayacağımız hiç bir rapor ya da araştırma belgesinde sizin kimliğinizi ortaya çıkaracak herhangi bir bilgiye yer verilmeyecektir.

Gönüllülük

Bu araştırmaya katılmak tamamen isteğe bağlıdır. Katıldığınız takdirde çalışmanın herhangi bir aşamasında herhangi bir sebep göstermeden onayınızı çekmek hakkına da sahipsiniz. Katılıp katılmama kararınız kurumunuzla olan ilişkilerinizi etkilemeyecektir.

Sorularınız için:

Bu formu imzalamadan önce, çalışmayla ilgili sorularınız varsa lütfen sorun. Daha sonra sorunuz olursa, araştırmayı yürüten Hanife ÇAKICI'ya cakic002@umn.edu e-mail adresinden ya da (90) 553 460-9613 nolu telefondan ulaşabilirsiniz. Araştırmayla ilgili haklarınız konusunda Minnesota Üniversitesi etik kuruluna su adresten danışabilirsiniz: Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota, 55455 USA, 001 (612) 625-1650.

Katilim Beyanı:

Ben, (katılımcının adı) ...., yukarıdaki metni okudum ve katılmam istenen çalışmanın kapsamını ve amacını, gönüllü olarak üzerime düşen sorumlulukları tamamen anladım. Çalışma hakkında soru sorma imkanı buldum. Bu çalışmayı istediğim zaman ve herhangi bir neden belirtmek zorunda kalmadan bırakabileceğimi ve bıraktığım takdirde herhangi bir olumsuzluk ile karşılaşmayacağımı anladım.

Bu koşullarda söz konusu araştırmaya kendi isteğimle, hiçbir baskı ve zorlama olmaksızın katılmayı kabul ediyorum.

Formun bir örneğini aldım / almak istemiyorum (bu durumda araştırmacı bu kopyayı saklar).

Katılımcının Adı-Soyadı:...
İmzası:...
Tarih (gün/ay/yıl):...../......./..............

Appendix C. Institutional Permission to Conduct Study at MoNE

T.C.
MİLLÎ EĞİTİM BAKANLIĞI
Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü

Sayı : 81576613/605/2237760   28/08/2013
Konu: Araştırma İzni

Sayın: Hanife ÇAKICI
(Ankara)

İlgi: 26.08.2013 tarihli dilekçe.

İlgi yazı ile Genel Müdürlüğümüze başvurmuş olduğunuz "İzleme/Değerlendirme Çalışmalarının Türkiye'de Eğitimde Karar Alma Sürecinde Algılanan Değeri" konulu doktora tezinizde kullanılacak veri toplama araçları ve araştırma iznine ait talebiniz Genel Müdürlüğümüzce incelenmiştir.

Onaylı bir örneği Bakanlığımızda muhafaza edilen, uygulama sırasında da mühürlü ve imzalı örnekten çoğaltılan 2 sayfa 15 sorudan oluşan Mülakat/Görüşme Sorularının gönüllülük esas olmak kaydıyla, Bakanlığımızın konuyla ilgili Merkez Birimlerinde uygulanmasında bir sakınca görülmemektedir.

Bilgilerinizi ve gereğini rica ederim.

Mustafa KOÇ
Bakan a.
Genel Müdür

EK: Veri Toplama Araçları (3 sayfa)

Güvenli Elektronik İmzalı
Aslı ile Aynıdır
2 9 Ağustos 2013

Appendix D. NVivo Coding Tree

1) **Educational planning/programming**
   a) Organizational structure
      i) Roles/responsibilities (each unit)
      ii) Collaboration / Integration across units
      iii) Sustainability of policymaking
         (1) Institutional memory
         (2) Stage
      iv) Bureaucratic culture
   b) Decision-making style
      i) Decentralized / Participatory
      ii) Centralized / Hierarchical
   c) Politics
      i) Party politics
      ii) Budgetary constraints
      iii) Change in top administration

2) **History of evaluation in educational planning/programming**
   a) National evaluations
      i) Governmental
      ii) Nongovernmental
   b) International evaluations
   c) Professionalism / Expertise in the field
   d) Turkish culture

3) **Institutionalization of M&E**
   a) Legal framework
      i) Rule of Law
      ii) Law 652
      iii) PFMC 5018
   b) Design
      i) Systems approach
         (1) M&E role and responsibility
         (2) Human Capital
         (3) Data management
      ii) Evaluation silos
4) **Concept of evaluation**
   a) Inspection and compliance
   b) Internal control and audit
   c) Performance management
   d) Monitoring
   e) Research
   f) Quantification

g) Competitiveness and materialism
h) Interpretive / Qualitative
i) Pilot studies
j) Procurement
k) Criticism
l) Justice

5) **Evaluand**
a) Projects
b) Pilot programs
c) Policies
d) Educational products
e) Sample studies

6) **Uses and purposes of evaluation**
a) Evidence-based (data driven) decision making
   i) Improvement / Learning
   ii) Accountability / Transparency
   iii) Policy development
b) Purpose
   i) Goal attainment / Outcome
   ii) Process
c) Marketing
   i) Right to information
   ii) Fear of criticism
   iii) Institutional image/reputation

7) **External suppliers of evaluation**
a) Academia
b) Civil society
c) International donors