TEACHER SELF-EFFICACY BELIEFS TOWARD MEASUREMENT AND EVALUATION PRACTICES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSTIY

BY

FATMA RANA CEYLANDAĞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN THE DEPARTMENT
OF
EDUCATIONAL SCIENCES

SEPTEMBER 2009

Approval of the Graduate School of Social Sciences

_____

Prof. Dr. Sencer Ayata
Director

I certify that thesis satisfies all the requirements as a thesis for the degree of Master of Science.

_____

Prof. Dr. Ali Yıldırım
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____

Assist. Prof. Dr. Yeşim Çapa Aydın
Supervisor

Examining Committee Members

Assoc. Prof. Dr. Oya Yerin Güneri (METU, EDS) _____

Assoc. Prof. Dr. Jale Çakıroğlu (METU, ELE) _____

Assist. Prof. Dr. Yeşim Çapa Aydın (METU, EDS)_____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Surname, Name: Ceylandağ, F. Rana

Signature :

ABSTRACT

TEACHER SELF-EFFICACY BELIEFS TOWARD MEASUREMENT AND EVALUATION PRACTICES

Ceylandağ, F. Rana

M.S., Department of Educational Sciences

Supervisor: Assist. Prof. Dr. Yeşim Çapa Aydın

September 2009, 94 pages

Teacher self-efficacy refers to teachers' belief in their abilities to perform an action. In the present study, a new scale was developed to measure teacher self-efficacy beliefs toward measurement and evaluation practices, called "Teacher Self-Efficacy toward Measurement and Evaluation Practices Scale" (TEMES). The purpose of this study was to test a model of relationships among teacher self-efficacy toward measurement and evaluation practices, teachers' sense of efficacy, year in teaching, and frequency of using traditional and alternative measurement and evaluation tools. Three hundred ninety-four teachers participated in the study. Confirmatory Factor Analysis (CFA), Multivariate Analysis of Variance (MANOVA), Canonical Correlation Analysis, and Structural Equation Modeling (SEM) were conducted to answer the research questions.

CFA provided evidence for five-factor structure of the TEMES. Cronbach's alpha coefficients of these five factors were satisfactory, ranging from .76 to .87. Teachers reported more frequent use of traditional measurement and evaluation tools than alternative tools. Separate MANOVAs yielded non-significant effect of gender on the factors of TEMES, but of teaching level. In addition, findings of canonical correlation analysis indicated that factors of TEMES were correlated with factors of Turkish teachers' sense of efficacy scale (TTSES). Results of the SEM indicated that teacher self-efficacy toward measurement and evaluation practices was positively correlated with frequency of using traditional and

alternative measurement and evaluation tools. Year of teaching was found to be a non-significant predictor of teachers' sense of efficacy, teacher self-efficacy toward measurement and evaluation practices, and frequency of using traditional and alternative measurement and evaluation tools.

Keywords: Self-efficacy, Teacher Self-efficacy, Measurement and Evaluation Practices

ÖZ

ÖLÇME-DEĞERLENDİRME UYGULAMALARINA YÖNELİK ÖĞRETMEN ÖZYETERLİĞİ

Ceylandağ, F. Rana
Yüksek Lisans, Eğitim Bilimleri Bölümü
Tez Yöneticisi: Yrd.Doç.Dr. Yeşim Çapa Aydın
Eylül 2009, 94 sayfa

Öğretmen özyeterliği, bir öğretmenin mesleğinin gerekliliklerini gerçekleştirmeye olan inancıdır. Bu çalışmada, ölçme-değerlendirme uygulamalarına yönelik öğretmen özyeterliğini ölçmek için yeni bir ölçek geliştirilmiş ve bu ölçek Ölçme-Değerlendirme Uygulamalarına Yönelik Öğretmen Özyeterliği Ölçeği olarak adlandırılmıştır. Çalışmanın amacı, ölçme-değerlendirme uygulamalarına yönelik öğretmen özyeterliği, genel öğretmen özyeterliği, meslekteki yıl, alternatif ve geleneksel ölçme-değerlendirme araçlarını kullanım sıklığı arasındaki ilişkiyi açıklayan bir model test etmektir. Çalışmaya 394 öğretmen katılmıştır. Araştırma sorularına cevap bulmak için Doğrulayıcı Faktör Analizi, Çoklu Varyans Analizi, Kanonik Korelasyon Analizi ve Yapısal Eşitlik Modeli (YEM) kullanılmıştır.

Doğrulayıcı Faktör Analizi, Ölçme Değerlendirme Uygulamalarına Yönelik Öğretmen Özyeterliği Ölçeği'nin 5 faktörlü yapıda olduğunu göstermiştir. Bu beş faktörün Cronbach alfa katsayıları tatmin edicidir ve .76 ile .87 arasında değişmektedir. Öğretmenler, alternatif ölçme değerlendirme araçlarını geleneksel ölçme-değerlendirme araçlarına göre daha sık kullandıklarını belirtmişlerdir. Çoklu Varyans Analizleri, yeni ölçeğin beş faktörü üzerindeki cinsiyet etkisinin istatistiksel olarak anlamlı olmadığını, fakat öğretim seviyesinin fark yarattığını ortaya çıkarmıştır. Ayrıca Kanonik Korelasyon Analizi sonuçları, yeni ölçek

faktörlerinin Öğretmen Özyeterlik Ölçeği'nin faktörleriyle ilişkili olduğunu göstermiştir. YEM analizinin sonuçları, ölçme ve değerlendirme uygulamalarına yönelik öğretmen özyeterliğinin alternatif ve geleneksel ölçme değerlendirme araçlarının kullanım sıklığı ile olumlu bir ilişkisi olduğuna işaret etmiştir. Ancak öğretmenlerin meslekte geçirdikleri yıl ile öğretmen özyeterliği, ölçme değerlendirmeye yönelik öğretmen özyeterliği, alternatif ölçme değerlendirme araçlarının kullanım sıklığı ve geleneksel ölçme değerlendirme araçlarının kullanım sıklığı arasında istatistiksel olarak anlamlı bir ilişki bulunamamıştır.

Anahtar Kelimeler: Özyeterlik, Öğretmen Özyeterliği, Ölçme ve Değerlendirme Uygulamaları

To my parents and lovely sister

ACKNOWLEDGMENTS

I am heartily thankful to my supervisor, Assist. Prof. Dr. Yesim Çapa Aydın for her guidance and support from the beginning to the end of this research. She was very generous and patient in answering my endless questions, identifying and correcting any gaps. Attending her graduate level courses and working with her during my thesis research was a pleasure for me.

I would like to thank Assoc. Prof. Dr. Oya Yerin Güneri and Assoc. Prof. Dr. Jale Çakıroğlu for their contribution to this study and the motivating discussions during the thesis defense.

I wish to express my deep gratitude to Assist. Prof. Dr. Finlay McQuade, who supported me in any respect during the time I had the opportunity to assist him in his work. He helped me broaden my perspective on my goals along with my assessment and problem solving skills.

My deepest thanks to my dearest sister, Gökce Girgin, and my brother-in-law, Murat Girgin, for their encouragement and warmth. I had the enormous pleasure of spending time with you in the United States, and you have given me motivation by your passion-filled speeches. Thank you for being there despite the distance.

I extend particular thanks to Özer Özaydın for his genuineness, faithfulness, unconditional love, endless support, and emphatic understanding.

I am also indebted to my friends Funda and Ayşegül, who have persistently urged me to stay in the library for long hours despite craving coffee and chocolate.

Special thanks to my dearest friend, Elif Sürer Köse for her sincerity and unconditional positive regard. Thank you for being a true friend, who is distant in miles but close at heart.

Finally, and most importantly, thanks to the giants of the statistics in social sciences, Barbara G. Tabachnick, Linda S. Fidell, Andy Field, Hair et al. and others whom I have not mentioned. Thank you for writing the thickest books of all times.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

TTKB: The Authority of Turkish Board of Education

MoNE: Turkish Ministry of National Education

ERDHO: Educational Research and Development Head Office

HSEC: METU Human Subjects Ethics Committee

TEMES: Teacher Self-Efficacy toward Measurement and Evaluation Practices Scale

FMES: Frequency of Using Different Measurement and Evaluation Tools Scale

Alternative-ME: Frequency of Using Alternative Measurement and Evaluation Tools

Traditional-ME: Frequency of Using Alternative Measurement and Evaluation Tools

TTSES: Turkish Teachers' Sense of Efficacy Scale

SPSS: Statistical Package for Social Sciences

EFA: Exploratory Factor Analysis

M: Mean

SD: Standard deviation

ANOVA: Analysis of Variance

MANOVA: Multivariate Analysis of Variance

AMOS: Analysis Moments of Structures

CFA: Confirmatory Factor Analysis

SEM: Structural Equation Modeling

NNFI: Non-normed Fit Index

CFI: Comparative Fit Index

RMSEA: Root Mean Square of Approximation

CHAPTER I

INTRODUCTION

In the following sections, the reason for researchers' decision to study on teachers' efficacy beliefs toward measurement and evaluation practices, the purpose and the significance of the study, and definition of the terms are reported in detail.

1.1. Background of the Study

Measurement and evaluation are important in terms of including the activities in which teachers can get information to modify or improve instructional strategies (Boston, 2002). If teachers know about students' progress and needs in learning, they can decide to try alternative methods, use additional materials to teach or persist how they teach. What if a teacher thinks that she or he is not good enough at assessing student learning and evaluating the results of assessment?

It has been suggested that there are problems in measurement and evaluation applications in public schools (Ministry of National Education, 2005, 2006). Moreover, most of the teachers suffer from not having enough background in using the techniques of student assessment proposed in the new educational program. Teachers also reported having difficulty in preparing and administering assessment tools, and making use of the results of student assessment (Gelbal & Kelecioğlu, 2007). In an extensive study conducted by the Turkish Ministry of National Education (MoNE) and Educational Research and Development Head Office (ERDHO), general teacher qualifications in different teaching activities

were examined, e.g., knowing student, developing instructional strategies, measurement and evaluation, communication with parents and other stakeholders. One of the striking results of this study was that the mean score of qualification ratings of teachers were the lowest in measurement and evaluation practices and communicating with parents or other teachers in the school among other areas (like use of instructional strategies, development of educational program, and content knowledge). In addition, participants also stated that they need help for developing their skills in using alternative assessment methods, analyzing the results of student assessment, and giving feedback to students and their parents about student evaluation. In the light of these results, the researchers concluded that teachers strongly need in-service training in measurement and evaluation practices and teachers' perception toward measurement and evaluation practices may change in a positive way by this support (MoNE & ERDHO, 2006).

These studies have led researchers conduct studies on teacher self-efficacy toward measurement and evaluation practices. In the study which was conducted by MoNE and ERDHO, it was stated that teachers were asked for their perception toward their qualifications in teaching. However, perception can occur under the effect of interacting factors, such as past experiences and culture (Chalmers, 1997). Since self-efficacy is a construct that differs from perception in a way that people question themselves only in a particular action, it can be practical and meaningful to examine teachers' efficacy beliefs rather than their perceptions toward measurement and evaluation practices.

1.2. Purpose of the Study

First of all, researchers intended to examine teacher self-efficacy toward measurement and evaluation practices. Since there is no instrument to measure teachers' efficacy beliefs toward measurement and evaluation practices, a new

scale was developed and validated in this study. During literature search, the researchers realized that year in teaching can be an important variable which can influence teacher self-efficacy toward measurement and evaluation practices. In addition, another variable, frequency of using different measurement and evaluation tools, was considered that can distinguish the teachers who are efficacious in measurement and evaluation practices from the teachers who are not.

All in all, there were two main purposes of this study: One was to develop an instrument to measure teacher self-efficacy toward measurement and evaluation practices and the other was to test a model of relationships among teacher self-efficacy toward measurement and evaluation practices, teachers' sense of efficacy, year in teaching, and frequency of using traditional and alternative measurement and evaluation tools.

1.3. Significance of the Study

Teacher self-efficacy is an issue which has been studied for almost 30 years and there have been many scales developed to assess teacher self-efficacy during these studies (Henson, 2002). It is also possible to see research studies examining the relationship between teacher self-efficacy and various variables such as student self-efficacy, student achievement, and teacher behavior. Further, many scales were developed to assess teacher self-efficacy in different fields such as classroom management, student engagement, and science teaching (Tschannen-Moran, Woolfolk Hoy & Hoy, 1998). On the other hand, efficacy items related to measurement and evaluation practices appear in small numbers (Karaca, 2008).

In one of the previous studies held in Turkey, Çakan (2004) reported that teachers perceive themselves inadequate in measurement and evaluation practices and most

of the teachers from various teaching grades prefer to use traditional methods of measurement and evaluation. Regarding the results of Çakan's study, developing an instrument which assesses teacher self-efficacy particularly in measurement and evaluation practices can make a contribution to what is known about teachers' efficacy beliefs about measurement and evaluation practices.

It has been proposed that as teachers gain experience in teaching, they may develop self-efficacy toward measurement and evaluation practices. Bandura (1997) also suggested that mastery experiences (own performances of people) is the most important source for developing self-efficacy. Thus, year in teaching was considered as an important variable in the present study. In addition to year in teaching, frequency of using different measurement and evaluation practices was included as another variable in this study to investigate the proposition that teachers who have higher self-efficacy tend to try new methods in measurement and evaluation. Similarly, Gibson and Dembo (1984) found that efficacious teachers are open to new ideas; therefore, in the present study it is expected that efficacious teachers may have a tendency to try alternative measurement and evaluation tools rather than traditional ones.

1.4. Definition of the Terms

Self-efficacy: Belief in one's capabilities to organize and execute the courses of action required to produce given attainments (Bandura, 1997, p. 3).

Teachers' sense of efficacy: Teacher's belief in his or her capability to organize and execute courses of action required to successfully accomplishing a specific teaching task in a particular context (Tschannen-Moran et al., 1998, p.22).

Teacher self-efficacy toward measurement and evaluation practices: Teacher's belief in his or her ability in measurement and evaluation practices.

CHAPTER II

REVIEW OF THE LITERATURE

In this chapter, theoretical framework for the study was represented with the leading studies on self-efficacy, teachers' sense of efficacy and measurement of self-efficacy beliefs. Firstly, the construct of self-efficacy is introduced under the framework of Social Cognitive Theory. This is followed by the section describing how self-efficacy belief was measured and the psychometric properties of the existing self-efficacy scales. Lastly, teachers' sense of efficacy is defined and measurement studies related to teacher self-efficacy in Turkey and other countries are presented in a chronological order.

2.1. Self-Efficacy

In his book named as *Self-efficacy: the Exercise of Control*, Bandura (1997) defined self-efficacy as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (p.3). The concept of self-efficacy arose from Bandura's Social Cognitive Theory in 1977. According to this theory, human behavior, environment and personal factors interact and influence each other through the process of reciprocal determinism (presented in Figure 2.1.) (Bandura, 1997). In this theory, reciprocal causality implies that there is a bidirectional interaction between personal factors, behavioral patterns and environmental influences. For example, a person's self-efficacy (personal factor) can be an indicator of how he or she self-regulates the performance (behavior), and their performance can affect their future self-efficacy beliefs in turn (Bandura, 1997).

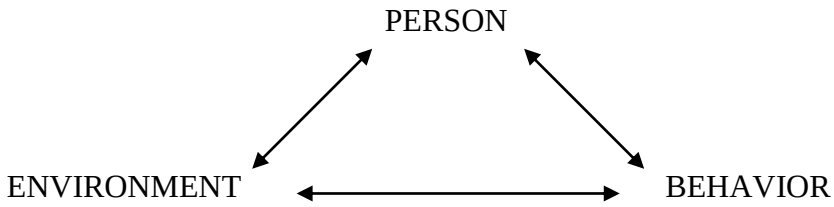

Figure 2.1 *Theoretical Model of Triadic Reciprocal Determinism*
Source: Bandura (1986, p. 24)

Bandura (1997) emphasized that perceived self-efficacy contributes to the acquisition of knowledge structures related to possessed skills by influencing motivation and the choice of activities. Therefore, perceived self-efficacy has an important role in Social Cognitive Theory. Bandura, Caprara, Barbaranelli, Gerbino, and Pastorelli (2003) found that people with high self efficacy tend to display the behavior of cooperativeness, helpfulness, sharing with others, and caring for others' welfare. The most important characteristic of self-efficacy is that self-efficacy is *task* and *situation* specific (Bandura, 1997). That is, self-efficacy beliefs may differ according to the task they are responsible for and the situation in which they perform. For example, one may feel comfortable with writing an essay but not with speaking in public.

In addition, Bandura (1997) reported that efficacy beliefs differ in *level, generality* and *strength*. People's self-efficacy may differ in level by the contribution of difficulty of task demands. An example for *level* is when the athletes are asked to judge their high-jumping efficacy; they would consider whether or not they can jump over the barriers at different heights. While mentioning the *generality* dimension, Bandura (1997) stated that people may think themselves as efficacious on either many of the activities or on just a few of them. Moreover, efficacy beliefs vary in *strength*; that is, having stronger sense of efficacy beliefs causes an increase in perseverance in the face of difficulties and hence possibility of being successful.

2.1.1. Four Sources of Self-Efficacy Beliefs

Bandura (1997) proposed that self-efficacy beliefs develop through four sources of influence. These sources are named as *enactive mastery experience* (which is the one that you perform by your own), v*icarious experiences* (those are the ones that you observe others' performance on a particular task), s*ocial persuasion* (being approved by someone who is professional in the area like a supervisor or a colleague) and, *physiological and emotional states* (e.g. physical accomplishments, health functioning, coping with stress).

Bandura (1997) noted that the most influential source of efficacy is enactive mastery experiences since they give the most realistic evidence of whether an individual can perform whatever it takes to succeed. If people succeed only in easy tasks, then they start to expect quick results and give up by failures. According to Bandura (1997), successful performances do not contribute to self-efficacy and failures do not lower self-efficacy all the time. The contribution to the development of self-efficacy depends on people's questioning their capability toward a success or failure. Further, mastery experiences will contribute to one's self-efficacy belief in consideration with *level, strength* and *generality* dimensions. While simple tasks may result in belief that they succeed only in easy tasks but not in difficult ones, tasks requiring perseverance will contribute much more to self-efficacy beliefs.

For vicarious experience, Bandura (1997) suggested that mastery experiences cannot be the only source of information about people's capabilities. Efficacy beliefs are influenced by experiences of other people, and these are named as vicarious experiences. When somebody sees, hears from others or gives evidence that others perform in any kind of task, especially the hard ones, he can start to believe that he may perform in the same task as well. Bandura (1997) gave the

example that high jumpers can compare their proficiency and their improvement with the previous heights reached by other athletes. He concluded that people assess their capability in comparison with their peers or colleagues.

Another source of efficacy judgments is verbal persuasion. If other people make someone believe that he or she is capable of doing something, it can be easier to struggle with difficulties in performing an action (Bandura, 1997). Therefore, people who are persuaded verbally and capable of performing an action will show greater effort, and keep on trying. Finally, affective states can have considerable influence on self-efficacy beliefs of people. In this respect, enhancing physical status, decreasing the effect of stress and emotional tendencies can be a way of developing positive self-efficacy beliefs (Bandura, 1991).

Considering Bandura's four sources of efficacy information, Gist and Mitchell (1992) suggested that there are major questions to ask when people judge their capabilities. These questions are: What do different tasks require? How much does an individual attribute a failure or achievement to himself or herself? How does each performance contribute to self-efficacy? Furthermore Gist and Mitchell (1992) proposed three strategies to change self-efficacy beliefs. These strategies include providing the individual information to understand the task attributes, providing the individual information about how efficacy beliefs develops (*i.e.*, the sources of self-efficacy), and providing the individual guideline about how much effort he or she should make to develop self-efficacy beliefs.

2.1.2. Self-Efficacy and Other "Self" Constructs

When self-efficacy is compared with other "self" concepts such as self-concept, self-confidence, self-esteem, and self-worth, self-efficacy differs from those in terms of being specific to a particular task (Tschannen-Moran, Woolfolk Hoy &

Hoy, 1998). Bandura (1997) stated how self-concept is measured and the difference between self-concept and self-efficacy. In Bandura's words, self-concept contributes "understanding of people's attitudes toward themselves and how these attitudes may affect their general outlook on life" (p. 11). In addition, Bandura (1997) stated that measurement of self-concept is done by asking people how many appreciable characteristics they attribute themselves. In the light of self-concept measurement studies, Bandura (1997) concluded that the predictive value of self-concept decreases when the influence of self-efficacy is considered in someone's personal belief.

Another similar concept, self-confidence is defined as believing in oneself (Benabou & Tirole, 2002). In that sense, this construct seems a general view of a person about himself and not an opinion about his characteristics specific to a situation as in self-efficacy.

While differentiating self-efficacy from self-esteem, Bandura (1997) reported that "perceived self-efficacy is concerned with judgments of personal capability; whereas self-esteem is concerned with judgments of self-worth." (p. 11). That is, self-esteem deals with how much an individual appreciates himself. However, the question of self-efficacy is how well people can act in different task situations. Similarly, Pajares (1996) also pointed out that self-esteem and self-efficacy differs from each other with the questions used to assess them. For example, "how I define myself" and "how I feel about myself" are the questions referring to self esteem; self-efficacy considers the ones like "how well can I solve this science problem?" or "how well can I write a bestseller book?"

Other than the "self" concepts discussed in the literature, one more distinction is needed to be made between self-efficacy and outcome expectancy, since both have a relationship with self-regulation. Gist and Mitchell (1992) reported that

"self-efficacy is one of several cognitive processes frequently considered in self-regulation." (p. 186). Self-efficacy was considered in a relationship with outcome expectancy which was defined as expectancy about consequences of a performance by Bandura (1997). People who are self-efficacious have a tendency to show more effort to attain their expectations, when they face with a difficulty in performing an action (Bandura & Cervone, 1986). However, self-efficacy differs from outcome expectancy in that it is a belief in one's ability to perform a particular action. For example, Zimmerman (2000) stated that a student's belief on getting grade A is a kind of self-efficacy belief, whereas considering this grade as a useful indicator to get a good job refers to outcome expectancy. In consideration with the definition of outcome expectancy, locus of control, whether people have the control of their behavior, should be defined at this point. According to Rotter (1966), locus of control is related to how people relate internal and external factors to their outcome. Internal locus of control refers to belief in self-responsibility for failure or success while external locus of control means that a person relate his failure or success to external factors, such as fate, luck, or external circumstances (Rotter, 1966). Bandura (1997) also reported that locus of control is an inconsistent predictor of different behaviors which can be uniquely explained by self-efficacy.

2.1.3 Measurement of Self-efficacy Beliefs

Self-efficacy is a construct that has attracted many researchers in social sciences and this led to the development of several instruments measuring this construct. There are many self-efficacy scales assessing people's self-efficacy in different fields such as alcohol resistance (Rychtarik, Prue, Rapp, & King, 1992), parenting (Bandura, Caprara, Barbaranelli, & Pastorelli, 2001), career decision (Betz, Klein & Taylor, 1996), teaching (Tschannen-Moran & Woolfolk Hoy, 2001), computer usage (Thatcher & Perrewé, 2002), and geometry (Cantürk-Günhan & Başer,

2007).

There are some points to consider while developing instruments to measure self-efficacy. According to Bandura (1997), there has been a discussion on what a scale measuring self-efficacy should question; should it ask for beliefs on performing an action but not the personal qualities? Later, Bandura (2006) reported guidelines which should be considered in development of a self-efficacy scale: First, the items of the instrument should include "can" or "will" as a judgment of capability and a statement of intention, respectively. This is because of the fact that self-efficacy is a judgment of how much a person can perform in a specific task (Bandura, 1997). Second, the scale should be unipolar. That is, the scale cannot include negative integers like -1, -2, -3 etc. Because zero value does not indicate any gradation, it is not recommended to use negative numbers (e.g., -1, -2) in the scale. Third, it should be guaranteed to the participants that their answers will not be shared with others. Otherwise, people would feel uncomfortable with others' judgment on their ideas about themselves. Lastly, it is very important to make self-efficacy scales have predictive validity; hence, self-efficacy interests people's future performance on a given task (Bandura, 2006).

2.2. Teachers' Sense of Efficacy

It is possible to derive the definition of teacher self-efficacy from the description of self-efficacy as "teacher's belief in his or her capability to organize and execute courses of action required to successfully accomplishing a specific teaching task in a particular context" (Tschannen-Moran et al., 1998, p.22). Some researchers defined teacher self-efficacy as teachers' beliefs in their abilities to affect student performance (Armor et al., 1976; Gibson & Dembo, 1984). In addition to affecting student performance, Dellinger, Bobbett, Olivier and Ellett (2007) emphasized that teacher self-efficacy focuses on outcome of successful teaching

behaviors and student characteristics and behaviors. In addition, Bandura (1997) pointed out that low teacher efficacy beliefs can give rise to low student efficacy and low academic achievement, and these may yield to negative teacher self-efficacy beliefs. Furthermore, teachers' sense of efficacy beliefs has a strong influence on not only student performance but also on how much goals are achieved, and how much a teacher changes (Tschannen-Moran et al., 1998).

According to Bandura (1994), self-efficacy beliefs have an impact on how people make their choices, on their level of motivation, their resilience against difficulties or stressors, and their sensitivity to depression. In that sense, it is not very hard to predict which factors would affect teacher self-efficacy. There are many research studies showing the relationship between student achievement and three kinds of efficacy which are students' self-efficacy, teacher self-efficacy and collective efficacy (Pajares, 1996; Tschannen-Moran et al., 1998). Gibson and Dembo (1984) reported that teachers who have high self-efficacy work longer with a student who has difficulty in learning. Moreover, teacher self-efficacy beliefs influence their resilience against the difficult situations (Gibson & Dembo, 1984). These results are supported by recent studies. For example, Ware and Kitsantas (2007) found that efficacious teachers display greater effort for teaching and feel responsible for both their failures and achievements.

2.2.1. Measurement of Teachers' Sense of Efficacy Beliefs

As well as some instruments were developed to measure teacher self-efficacy in teaching a subject area such as efficacy in science teaching (Riggs & Enochs, 1990), efficacy in computer teaching (Akkoyunlu, Orhan, & Umay, 2005), efficacy in geography teaching (Karadeniz, 2005), it is possible to notice that some scales on teacher self-efficacy included the factors on personal teaching efficacy and general teaching efficacy (Gibson & Dembo, 1984); efficacy to

influence decision making, school resources, instruction, discipline, efficacy to enlist parental involvement, community involvement, and efficacy to create a positive school climate (Bandura, 2001); teacher self-efficacy in classroom management, instructional strategies, and student engagement (Tschannen-Moran & Woolfolk Hoy, 2001).

Measurement studies of teachers' sense of efficacy beliefs have started by the research of RAND organization on student learning and teachers' characteristics in 1976. There were just two items which could be identified classified as measuring teachers' self-efficacy. However, this study shed light to other studies measuring what teachers' opinion was on their personal responsibility in student learning (Guskey & Passaro, 1994).

RAND Items (1976). The first example of assessing teacher efficacy was observed in the study of Rand Corporation in 1976. The main purpose of the study was to increase reading scores of elementary students by defining most successful school and classroom policies and other variables (Armor et al., 1976). To determine those, the researchers examined the success of different reading programs and interventions. There were two items measuring teacher efficacy and these two focused on how teachers may influence student motivation (Tschannen-Moran et al., 1998). In this study, researchers concluded that teacher efficacy was one of the significant factors that had an influence on reading achievement of elementary students (Armor et al., 1976).

Rose and Medway (1981). The relationship between teacher's locus of control and student learning was examined in this study. Locus of control was defined in a preceding study of Rotter (1966). According to Rotter (1966) locus of control is related to how people relate internal and external factors to their outcome. Internal locus of control refers to belief in self responsibility for failure or success while

external locus of control means that a person relate his failure or success to external factors, such as fate, luck, or external circumstances (Rotter, 1966). Rose and Medway (1981) found significant relationship between teachers' locus of control and student achievement.

Webb Scale (1982). This scale was developed in order to contribute to the measurement of teacher efficacy by expanding Rand's measure. In order to make participants avoid giving responses fitting social desirability, Webb and his colleagues used a forced-response format. Any reliability value or validation study has not been reported by the researchers (Tschannen-Moran et al., 2001).

Aston Vignettes (1984). Ashton, Buhr and Crocker (1984) developed a scale including vignettes describing situations a teacher acts and questions on how effective a teacher would be in that kind of situation. The scale had two versions in response as self-referenced with "extremely ineffective" to "extremely effective," and norm-referenced with "much less effective than most teachers" to "much more effective than other teachers." However, the instrument has not been accepted and used widely in the field.

Gibson and Dembo (1984). Gibson and Dembo (1984) stated that teacher self-efficacy beliefs are teachers' evaluation on how much they are able to create positive student change. In this concern, they developed a 30- item teacher self-efficacy instrument which included two factors named as *personal teaching efficacy* (PTE, alpha= .75) and *teaching efficacy* (GTE, alpha= .79). Gibson and Dembo (1984) concluded that validation studies are needed to stabilize the factor structure. After development of this instrument, there have been many research studies done on teacher self-efficacy and its relationship with teachers' classroom behaviors, openness to new ideas, and attitudes toward teaching.

Riggs and Enochs (1990). Another important study to measure teacher self-efficacy belief was done by Riggs and Enochs in 1990. They developed a 25-item instrument called *Science Teaching Efficacy Belief Instrument* (STEBI) to measure classroom teacher self-efficacy beliefs toward science teaching. This instrument included two factors named as *personal science teaching efficacy belief (*alpha= .92) and *science teaching outcome expectancy (*alpha= .77). Riggs and Enochs (1990) reported that their scale produces valid and reliable scores indicating teachers' belief toward science teaching and learning.

Bandura (2001). Bandura developed a teacher self-efficacy scale which included 30 items on a nine-point scale with seven subscales: efficacy to influence decision making, efficacy to influence school resources, instructional efficacy, disciplinary efficacy, efficacy to enlist parental involvement, efficacy to enlist community involvement, and efficacy to create a positive school climate. However, Bandura has not reported any finding regarding validity or reliability for his instrument.

Tschannen-Moran and Woolfolk Hoy (2001). Tschannen-Moran and Woolfolk Hoy (2001) reported that most of the teacher self-efficacy scales did not include items on personal competence and tasks which exist in teaching process. Moreover, Tschannen-Moran, Woolfolk-Hoy and Hoy (1998) argued the necessity of a valid and reliable teacher self-efficacy scale. In the light of these arguments, Tschannen-Moran and Woolfolk Hoy (2001) developed a new scale with 52 items and named it as *Teachers' Sense of Efficacy Scale (TSES)*, originally known as *Ohio State Teacher Self-Efficacy Scale (OSTES)*. To validate the scores obtained from this scale, Tschannen-Moran and Woolfolk Hoy (2001) constructed three different studies with 624 participants including pre-service and in-service teachers. At the end of these studies resulting scale had 24 items in the long form, and 12 items in the short form. To make sure that both two versions of the scale provide evidence for construct validity, Tschannen-Moran and

Woolfolk Hoy (2001) checked for the correlation between their scales and previously developed teacher self-efficacy scales as RAND items and Hoy and Woolfolk (1993)'s 10-item adaptation of Gibson and Dembo TES. Among the resulting correlation coefficients, the highest ones were obtained with the scale measuring personal teaching efficacy. To indicate that both forms of TSES measured the same construct, Tschannen-Moran and Woolfolk Hoy (2001) reported that the intercorrelations between short and long form of TSES were in between .95 and .98. Moreover, they conducted Principal-Axis Factoring with Varimax Rotation and concluded that TSES had a three-factor structure. The factors were named as *efficacy for student engagement* (ESE), *efficacy for instructional strategies* (EIS), and *efficacy for classroom management* (ECM). Reliability analysis indicated that total scale reliability was .94 and those three subscales had high Cronbach Alpha Coefficients as .87 for ESE, .91 for EIS, and .90 for ECM (Tschannen-Moran & Woolfolk Hoy, 2001). The alpha values and the validation study indicated that *Teachers' Sense of Efficacy Scale* was a valid and reliable measure to assess teachers' sense of efficacy in student engagement, instructional practices and classroom management (Tschannen-Moran & Woolfolk Hoy, 2001).

Schmitz and Schwarzer (2005). Based on Bandura's Social Cognitive Theory Schmitz and Schwarzer (2005) developed a 4-point response scale composing of 27 items and administered their scale to 300 German teachers. They reported the values .67, .76 and .65 for test–retest reliability of the instrument in three year study. Further, the scale was reported to be related with personal attitudes than general self-efficacy scale and this situation was emphasized as an evidence for discriminant validity.

Dellinger, Bobbett, Olivier and Ellett (2007). The latest measure of teacher self-efficacy beliefs was developed by Dellinger, Bobbett, Olivier and Ellett and

named as Teachers' Efficacy Beliefs System—Self Form (TEBS-Self). The scale was on a 4-point rating scale [weak belief in my capabilities (1), moderate beliefs belief in my capabilities (2), strong belief in my capabilities (3), and very strong belief in my capabilities (4)] composing of 30 items. This scale was used in three distinct studies of the researchers and they did not reach a consensus in terms of the factor structure of the scale (Dellinger et al., 2007).

Aforementioned instruments are summarized in Table 2.1.

Authors	Sample Items	Type of Rating Scale	Number of items in the scale
Armor et al. (1976)	If I really try hard, I can get through to even the most difficult or unmotivated students.	5-point Likert Scale	2 items on teacher self-efficacy
Rose & Medway (1981)	When the grades of your students improve, it is more likely a. because you found ways to motivate the students, or b. because the students were trying harder to do well.	A forced-choice format	28
Ashton et al. (1982)	A teacher should not be expected to reach every child; some students are not going to make academic progress.	A forced-choice format	7
Ashton et al. (1984) (Ashton Vignettes)	Your school district has adopted a self-paced instructional program for remedial students in your area. How effective would you are in keeping a group of remedial students on task and engaged in meaningful learning while using these materials?	5-point Liker t scale	50
Gibson & Dembo (1984)	If a student masters a new math concept quickly, this might be because I knew the necessary steps in teaching that concept.	6-point Likert scale	30
Riggs & Enochs (1990)	I understand science concepts well enough to be effective in teaching elementary science.	5-point Likert scale	25
Bandura (2001)	How much can you do to get children to follow classroom rules?	9-point Likert scale	30
Tschannen-Moran & Woolfolk Hoy (2001)	To what extent can you craft good questions for your students?	9-point Likert scale	24
Schmitz and Schwarzer (2005)	Even if I get disrupted while teaching, I am confident that I can maintain my composure and continue to teach well.	4-point Likert Scale	27
Dellinger, Bobbett, Olivier & Ellett (2007)	1. Weak belief in my capabilities. 2. Moderate belief in my capabilities. 3. Strong belief in my capabilities. 4. Very strong belief in my capabilities. Effective manage routine and procedures for learning tasks...	4-point response scale	31

Table 2.1
Items from Some Teacher Self-Efficacy Scales

2.2.2. Measurement Studies of Teachers' Sense of Efficacy Beliefs in Turkey

In Turkey, history of the studies on teacher self-efficacy is not very old beginning in 2000s. The researchers mostly adapted previously established instruments in their studies. The examples of instrument adaptation studies are the ones that belong to Yılmaz, Köseoğlu, Gerçek and Soran (2004), Bıkmaz (2004), and Çapa, Çakıroğlu, and Sarıkaya (2005). Further, Erdem and Demirel (2007), Akkoyunlu, Umay and Orhan (2005), Karadeniz (2005), and Karaca (2008) conducted the development and validation studies of instruments assessing teacher self-efficacy in different fields.

Yılmaz, Köseoğlu, Gerçek, and Soran (2004). Yılmaz et al. adapted the *Teacher Self-Efficacy Scale,* which was developed by Schmitz and Schwarzer in 2000 in Germany. In this study, the researchers translated the original survey and reported reliability and validity findings after administering the instrument to Turkish teachers. Yılmaz and his colleagues (2004) reported that the reliability of the adapted scale was found .79, as Cronbach alpha value. Moreover, they found two factors and decided on keeping eight items, whereas the original scale included 10 items. The factors of the adapted instrument were *coping behavior* (başa çıkma davranışı) and *reformist behavior* (yenilikçi davranış).

Bıkmaz (2004). Bıkmaz adapted the *Science Teaching Efficacy Belief Instrument* (STEBI) developed by Riggs and Enochs on teacher self-efficacy beliefs toward science teaching. In this study, the purpose of the researcher was to provide evidence for validity and reliability of the scale for classroom teachers in Turkey. Bıkmaz (2004) reported that the adapted instrument has two factors including 20 items. Cronbach's alpha coefficient for the first factor which was named as *self-efficacy belief* was .78, and for the second factor, *outcome expectancy,* it was .60. In addition, .71 was the reliability coefficient for the whole instrument.

Çapa, Çakıroğlu, and Sarıkaya (2005). Çapa and her colleagues (2005) stated that a valid measure for efficacy beliefs of teachers has not been developed in Turkey. In that sense, Çapa et al. (2005) adapted the *Teachers' Sense of Efficacy Scale (TSES)* which was developed by Tschannen-Moran and Woolfolk Hoy in 2001. The purpose of the study was to adapt TSES in Turkish, examine reliability values for subscales and the whole scale, and provide construct related evidence for the adapted version of TSES. Çapa, Çakıroğlu, and Sarıkaya (2005) ran Confirmatory Factor and Rasch analyses to examine the factor structure and to report reliability coefficients of the factors. The analyses resulted in reliability indices as follows: .82 for the first factor, *student engagement,* .86 for the second factor, *instructional strategies,* and .84 for the third factor, *classroom management*. Çapa et al. (2005) confirmed the three-dimensional structure of the Turkish Teachers' Sense of Efficacy Scale (TTSES) using the data of 628 Turkish pre-service teachers.

Akkoyunlu, Orhan, and Umay (2005). Akkoyunlu et al. developed a teacher self-efficacy scale for computer teachers in 2005. Before developing the instrument, Akkoyunlu and her colleagues (2005) asked ten different experts who were instructors in Faculty of Education of Hacettepe University and ensured that the instrument had the content validity. The latest version of the instrument was a 5-point Likert scale consisting of 12 items and it was named as *Teacher Self-Efficacy Scale for Computer Teachers* (Bilgisayar Öğretmenliği Özyeterlik Ölçeği). The data were collected from 315 senior students in computer education and instructional technologies departments of eight different universities in Turkey. Findings yielded one dimension. The alpha coefficient of the instrument was very high with a value of .93.

Karadeniz (2005). The instrument assessing teacher efficacy in teaching geography was established by Karadeniz (2005). She developed a self-efficacy scale of geography for pre-service teachers of social sciences. The developed

instrument had 19 items and these items were collected under three factors. The factors and the reliability alpha values were reported as follows: .86 for *transform geography knowledge into life skills* (coğrafyayı yaşam becerilerine dönüştürebilme), .76 for *self-efficacy beliefs* (yeterlik algısı), and .63 for *awareness of behaviors in geography* (coğrafya alanındaki davranışlarda farkındalık). In addition, the split half reliability coefficient was reported as .79.

Erdem and Demirel (2007). A new instrument was developed and validated to assess pre-service teachers' self-efficacy beliefs toward teaching by Erdem and Demirel in 2004. They studied with 346 student teachers attending six different departments of a faculty of education. The instrument was established as a 5-point Likert scale in a single-factor model and the reliability coefficient for the whole scale was reported as .92.

Karaca (2008). In this study, the purpose was to investigate the perceptions of primary and high school teachers toward measurement and evaluation in education in Turkey. To measure the perceptions of teachers toward assessment practices, Karaca (2003) constructed a 5-point Likert scale with 75 items. Actually, it was reported that teachers' perception levels of *efficacy* was proposed to be measured. However, items did not have one of the important properties that an efficacy scale should have like did not include "can" or "will" as a judgment of capability and a statement of intention (Bandura, 2006). Karaca (2008) collected the data from 225 primary and high school teachers who worked in Eskisehir, Turkey. According to the results of this study, independent samples t-test yielded non-significant difference between male and female teachers' perception levels of efficacy toward measurement and evaluation practices. In addition, it was found that high school teachers' perception levels of efficacy were found out to be higher than primary teachers' by independent samples t-test. The results of one way ANOVA indicated no significant difference in teachers' perception levels of

efficacy toward measurement and evaluation practices according to year in teaching.

Sample items from the instruments which were adapted and developed in these studies are summarized in Table 2.2.

Developers	Researchers who adapted the instrument to Turkish	Sample Items	Type of Rating Scale
Schmitz & Schwarzer (2000)	Yılmaz, Köseoğlu, Gerçek & Soran (2004)	Zor durumlarda bile ebeveynlerle iyi bir iletişim kurabilirim.	4-point Likert scale
Riggs & Enochs (1990)	Bıkmaz (2004)	Öğrencilerin fen dersindeki başarılarından öğretmen sorumludur.	5-point Likert scale
Tschannen-Moran and Hoy (2001)	Çapa, Çakıroğlu and Sarıkaya (2005)	Öğrencileri okulda başarılı olabileceklerine inandırmayı ne kadar sağlayabilirsiniz?	9-point scale
Akkoyunlu, Orhan & Umay (2005)	Akkoyunlu, Orhan & Umay (2005)	No item was reported	5-point Likert scale
Karadeniz (2005)	Karadeniz (2005)	Coğrafya konularına yönelik grafik ve tabloları yorumlayabilirim.	5-point Likert scale
Erdem & Demirel (2007)	Erdem & Demirel (2007)	I can ensure my students trust me by expressing my ideas and behaviors clearly.	5-point Likert scale
Karaca (2008)	Karaca (2008)	Öğretim hedeflerine ve hedef davranışlara uygun ölçme araçlarını belirleyebilme. Her bir maddenin ayırt ediciliğini hesaplayabilme.	5-point Likert scale

Table 2. 2
Items from Some Teacher Self-Efficacy Scales Adapted or Developed in Turkey

2.2.3. Research on the Relationship between Teachers' Sense of Efficacy and Other Variables

The relationship between teacher self-efficacy and many different variables such as commitment to teaching, developing instructional strategies, classroom management, student achievement, and motivation was studied in various research studies (Tschannen-Moran & Woolfolk Hoy, 2001). For example, teacher self-efficacy was found in a relationship with student achievement (Ross, 1992), planning and organization in teaching (Freidman & Kass, 2002), enthusiasm for teaching (Guskey, 1984), and meeting needs of students (Guskey, 1988).

2.2.3.1. The Relationship between Teacher Self-Efficacy and Year in Teaching

Teacher self-efficacy was found in a relationship with year in teaching (Hoy & Woolfolk Hoy, 1993), grade level (Çapa, 2005), teaching area of specialization (Ross, Cousins, Gadalla & Hannay, 1999), education level (Friedman, 2003), and student achievement (Lee, Dedrick & Smith, 1991). Among these variables, increase in year in teaching was found to have an impact on developing positive teaching efficacy in the study of Hoy and Woolfolk Hoy (1993). However, some researchers concluded that teacher self-efficacy decreased by increasing year in teaching experience (Dembo & Gibson, 1985; Ghaith & Yaghi, 1997). There were other studies showing differences in teacher efficacy among the teachers who have varying levels of teaching experiences. For example, year in teaching was reported as positively correlated to teacher self-efficacy in the study of Tschannen-Moran and Woolfolk Hoy (2007). In addition, Tschannen-Moran et al. (1998) suggested that self-efficacy beliefs of expert teacher are resistant to change. In the line with this suggestion, Woolfolk Hoy and Burke-Spero (2005)

reported that self-efficacy is more changeable in the early years of teaching. Furthermore, they reported that novice teachers who have positive self-efficacy beliefs develop positive attitude toward teaching and have less stress in their job in their first year of teaching. On the contrary, Karaca (2008) reported that teachers' perceptions of efficacy toward measurement and evaluation practices do not differ significantly by the change in years of teaching. Çakan (2004) found a similar result that experienced teachers' perceptions toward their qualification levels are not different than the novice teachers' perceptions. In this context, it is important to understand what influences teacher self-efficacy and which factors are affected by teacher self-efficacy by the changing years of teaching experience. In the present study, to clarify the relationship between year in teaching and teacher self-efficacy toward measurement and evaluation practices, the researchers examined whether teacher self-efficacy toward measurement and evaluation practices is correlated with year in teaching, and whether these relationships are in positive or negative direction.

2.2.3.2. The Relationship between Teacher Self-Efficacy and Frequency of Using Different Measurement and Evaluation Tools

Regarding the inference of Gibson and Dembo (1984) that efficacious teachers tend to be open to try new methods and are not against alternative methods in teaching, using different measurement and evaluation tools are supposed to be a characteristic of teachers who have positive self-efficacy in teaching. In addition, Vitali (1993) reported that efficacious teachers prefer performance-based assessment, which is a kind of alternative assessment method, rather than traditional tests. Similar results were also found by Ross, Cousins and Gadalla in 1996. Ross and his colleagues (1996) examined whether the effect of different teaching tasks on teacher self-efficacy was moderated by between teacher variables (i.e., subject, experience, gender, preference for student centered

instruction and alternative assessment techniques). Ross et al. (1996) clarified different teaching tasks as *feelings of past success, feelings of being well-prepared,* and *student engagement*. The conclusion of this study was that when perceived success was positively correlated to teacher self-efficacy, teachers tended to use traditional assessment techniques more. Teachers prefer alternative assessment techniques when teacher self-efficacy was related to feelings of preparedness. Ross and his colleagues (1996) attributed using alternative assessment techniques to teachers' ability to take risks and try new methods. Correspondingly, the finding of Gibson and Dembo (1984) about efficacious teachers' tendency to being openness to new methods supports the view of Ross and his colleagues (1996).

2.3. Summary of Related Studies

In previous sections, the definition of self-efficacy, the sources contributing to self-efficacy development, the definition of teacher self-efficacy and measurement studies on teacher self-efficacy and related factors were reported in a chronological order. In this way, researchers clarified when teacher self-efficacy was started to be considered as an important construct, how teachers' sense of efficacy was measured and which constructs or variables were thought to be related to it.

Related literature indicated that there was a relationship between year in teaching and teaching efficacy (Dembo & Gibson, 1985; Hoy & Woolfolk, 1993; Ghaith & Yaghi, 1997; Tschannen-Moran & Woolfolk Hoy, 2007). In addition to relationship, more change is possible in teaching efficacy in the early years of teaching according to Woolfolk Hoy and Burke-Spero (2005). They concluded that efficacious novice teachers tend to develop positive attitude toward teaching and have less trouble in the first year of teaching. However, Çakan's (2004)

finding that teachers' perception about their qualification levels had no correlation to year in teaching is a contradictory result to these findings. Karaca (2008) supported this result by reporting non significant relationship between teachers' perception levels of efficacy in measurement and evaluation practices and year in teaching. This contradiction in the literature findings encouraged researchers to conduct a study to examine the relationship between year in teaching and teacher self-efficacy toward measurement and evaluation practices.

Because efficacious teachers were found to take risks in teaching (Gibson & Dembo, 1984), they were expected to develop and administer alternative teaching methods without hesitation (Ross et al., 1996). In that sense, the researchers intended to investigate whether teachers who have positive self-efficacy toward measurement and evaluation practices have a tendency to prefer alternative measurement and evaluation tools to traditional ones.

CHAPTER III

METHOD

This chapter presents the research methodology of the study. In detail, research design, research questions, description of variables, participants' demographic information, and instruments used in the study are mentioned respectively. The last section introduces the data analysis employed in this study.

3.1. Research Design

This study was an associational research since the relationship between years in teaching, frequency of using different kinds of measurement tools and teachers' efficacy beliefs toward measurement and evaluation tools were examined. In associational research, relationships among two or more variables are investigated without manipulating variables. Moreover, numerical representation is possible to display the relationship between variables (Fraenkel & Wallen, 2008).

To measure teachers' efficacy beliefs toward measurement and evaluation practices, a 9-point scale with 24 items was developed. Necessary permissions to administer the instrument were taken from the METU Human Subjects Ethics Committee (HSEC) and Educational Research and Development Head Office (ERDHO) in Ankara. Data were collected from 394 experienced teachers who worked in public primary and high schools in Ankara, Samsun, and Istanbul. Data were collected between May and June of 2008.

3.2. Research Question

In order to measure teacher self-efficacy toward measurement and evaluation practices, an instrument was developed. By using this instrument, a model was tested in which the following main and sub-research questions were addressed:

What is the best model explaining the relationship between teacher self-efficacy in measurement and evaluation practices, years of teaching experience, teachers' sense of efficacy, and frequency of using alternative and traditional measurement and evaluation tools?

1. How well do years of teaching experience predict frequency of using alternative and traditional measurement and evaluation tools?

2. How well do years of teaching experience and teachers' sense of efficacy predict the teacher self-efficacy in measurement and evaluation practices?

3. How well does teacher self-efficacy in measurement and evaluation practices predict frequency of using alternative and traditional measurement and evaluation tools?

3.3. Description of Variables

This section provides the operational definitions of variables investigated in this study:

Years in teaching: This independent variable corresponds to the number of years the participant teacher has been teaching. It was a continuous variable and the level of measurement was considered as ratio.

Teachers' sense of efficacy: Mean score was computed for the Turkish Teachers' Sense of Efficacy Scale (TTSES). High score indicates high teachers' sense of efficacy. The level of measurement for this variable was considered as interval.

Frequency of using different measurement and evaluation tools: This variable of the study was measured on a 5-point rating scale (1 referred to "never" and 5 referred to "always") and scores were obtained out of 5 by taking mean of 17 items. Items were generated from the measurement and evaluation tools that the Turkish Ministry of National Education (MoNE) proposed in latest curriculum (Erdoğan, 2007). To examine whether efficacious teachers prefer more alternative or traditional methods, the researchers divided this variable into two distinct variables as frequency of using alternative and traditional measurement and evaluation tools, i.e., *Alternative-ME* and *Traditional-ME*, respectively. *Alternative-ME* was measured by 10 items and Traditional-ME was represented by 7 items.

Teacher self-efficacy beliefs toward measurement and evaluation practices: The dependent variable, assessing teachers' beliefs in their abilities to perform tasks related to measurement and evaluation practices, was measured by an instrument developed by the researchers. It included 24 items on a 9-point rating scale ranging from "nothing" (1) to "a great deal" (9). The mean score of each participant was generated out of 9. The level of measurement for this variable was considered as interval.

3.4. Participants

Target population of the study was the public school teachers who were working in elementary and secondary schools in three different cities of Turkey: Ankara (the districts of Çankaya and Sincan), Samsun (Center district), and İstanbul (the districts of Zeytinburnu, Bakırköy and Eyüp). Convenient sampling procedure

was performed within this target population. The cities preferred to collect data in this study were selected from three different regions of Turkey, because these are the ones convenient to the researchers. Data were collected from 44 elementary and secondary schools. The percentage of secondary schools was 47 and the rest (53%) were elementary schools. Table 3.1 displays the participating teachers' background data on gender, teaching level, branch, and graduation history.

Three hundred and ninety-four teachers participated in the study and these teachers were from public elementary and secondary schools. 57.11% of the participants were female and 42.89% of them were male. Participants' ages ranged from 22 to 63 and had a mean of 40. Year in teaching ranged from 1 to 40 with an average of 16. The percentage of teachers working in elementary schools was 53.05 and in secondary school were 46.95%. Twenty two percent of participating teachers had a science (i.e.,, teaching Physics, Chemistry etc.) and mathematics major, while 78% of them had a social science major (i.e.,, teaching Turkish, English, and Geography etc.).

Among these teachers, 4.3% of them were graduated from a teacher school, 11.7% of them were graduated from a pre undergraduate program (two-year university program), 77.9% of them had a bachelor's degree, and 6.1 % of them had a master's degree or Ph.D. degree. Approximately fifty-nine percent (58.9%) of all participants graduated from a faculty of education, whereas 41.1% of them graduated from other faculties rather than education faculty. The percentage of the ones who have taken a course on measurement and evaluation during university education was 86.5 and 13.5% of all participants have never taken a course on this issue. Lastly, 35.3% of all participants have joined an in-service training program, while 64.7% of participant teachers did not join such a training program (Table 3.1).

Table 3.1

Demographic Information of the Participants

	Percentage	N
Gender		
Female	57.11	225
Male	42.89	169
Teaching Level		
Elementary	53.05	209
Secondary	46.95	185
Branch		
Science	22	87
Social Science	78	307
Graduation		
Teacher School	4.3	17
Pre undergraduate	11.7	46
Undergraduate	77.9	307
Graduate	6.1	24
Faculty of Education		
Yes	58.9	232
No	41.1	162
Course		
Yes	86.5	341
No	13.5	53
In service Training		
Yes	35.3	139
No	64.7	255

3.5. Data Collection Instruments

Data were collected with an instrument composing of four sections: The first section was composed of the demographic information.

Section II included a scale of Teacher Self-Efficacy toward Measurement and Evaluation Practices (TEMES), which was developed by the researchers. The questionnaire was a 9-point scale ranging from "nothing" to "a great deal." The scale included the items generated from the teaching qualifications in measurement and evaluation practices which were developed by MoNE and ERDHO. The scale development procedure is presented in detail in section 3.5.2.

Section III included Frequency of Using Different Measurement and Evaluation Tools Scale (FMES), and it was developed by the researchers as a 5-point Likert scale including 17 measurement and evaluation tools suggested by the Turkish Ministry of National Education (MoNE) in the latest curriculum (Erdoğan, 2007). This scale was developed to measure the frequency of using different measurement and evaluation tools. Two different variables were extracted from this scale to measure frequency of using alternative and traditional methods and these were named as *Alternative-ME* and *Traditional-ME*. Alternative-ME, namely frequency of using alternative measurement and evaluation tools, was measured by ten items, and 7 items assessed Traditional-ME or frequency of using traditional measurement and evaluation practices. The score for these two variables were computed by adding the item scores and taking an average of total score dividing by the number of items. For example, mean score of Alternative-ME is equal to the total score of ten items divided by ten. Therefore, both Alternative-ME and Traditional-ME corresponded to a score out of five.

Section IV included Turkish Teachers' Sense of Efficacy Scale (TTSES). The scale was originally developed by Tschannen-Moran and Woolfolk Hoy in 2001

and was adapted to Turkish by Çapa, Çakıroğlu, and Sarıkaya (2005). The items include "how well can you…?" and "how much can you…?" patterns to meet the criteria of Bandura (2005) which are considered in developing self-efficacy scales. TTSES includes 24 items on a 9-point scale ranging from (1) "nothing" to (9) "a great deal." and these items measure teacher self-efficacy beliefs in three domains: classroom management, instructional strategies, and student engagement.

3.5.1. Demographic Information

In the original instrument, after the information about the purpose of the study and confidentiality of the results were stated, eleven questions were included in the demographic information section to determine the characteristics of the participating teachers in detail. In demographic information part, the categorical variables were gender, participating teachers' graduate degree (levels were teacher school, pre undergraduate, undergraduate, graduate, and doctorate), teaching branch (with levels of science and social science), teaching level (primary and secondary), school type (levels were public primary and public high school), whether they have taken any course on measurement and evaluation during the undergraduate education (levels were yes and no) and whether they have taken any in-service training toward measurement and evaluation (levels were yes and no). Age and year in teaching were continuous variables. In addition to these, the name of the faculty and the program which they were graduated from were asked as open ended questions and these were coded as one variable with two levels: being a graduate of a faculty of education or not.

3.5.2. Teacher Self-Efficacy toward Measurement and Evaluation Practices Scale (TEMES)

In order to examine how efficacious teachers are in consideration with measurement and evaluation practices, the researchers decided to develop a new scale in the light of the qualifications in teaching which the Turkish Ministry of National Education submitted in 2007. Before the item construction, resources on measurement of self-efficacy, available teacher self-efficacy scales (e.g., teachers' general efficacy, teachers' efficacy toward mathematics and science teaching), validity and reliability evidences for these scales were examined in detail.

3.5.2.1. Instrument Development

During the development of the instrument, the following steps were followed: deciding the dimension of the proposed instrument, generating items from different sources including the qualifications that Turkish Ministry of National Education proposed, determining the rating scale of the instrument reviewing items by experts, validating the items, administering the items to a development sample (i.e., conducting the pilot study), evaluating the items and deciding on the length of the scale (DeVellis, 2003, p. 60-100).

An item pool was generated considering the literature in this field. The primary source was the report on qualifications in teaching generated by the Turkish Ministry of National Education (MoNE) and Educational Research and Development Head Office (ERDHO). Under the sub-heading of *Observing Student Development and Evaluation,* there are 24 qualifications. These qualifications were written in question format starting with the pattern of "how much can you…?" or "how well can you…?" In addition to these, 9 more items were constructed in consideration by examining preceding teacher efficacy scales. During 2007 fall semester, the draft scale was reviewed by graduate students of

Test Construction course in Middle East Technical University and by five experts from educational sciences, elementary education, and measurement and evaluation departments of Middle East Technical University and Hacettepe University. They mostly focused on wording of the items and made some contributions on how the items may be revised to become more clear and understandable. In fact, review of the experts contributed to content validity of the instrument in terms of agreement on the content to be covered to measure the intended construct, which is teacher self-efficacy toward measurement and evaluation practices. The suggestions of the experts let the researchers decrease the number of items from 33 to 24 because there were some items related to each other and these items seemed redundant measuring the same construct.

After the items were generated, the rating scale was decided as a 9-point ranging from "nothing" to "a great deal." The reason of selecting a 9-point scale was Bandura's "Guide for Constructing Self-efficacy Scales." According to Bandura (2001, p. 7), "People usually avoid the extreme positions so a scale with only few steps may, in actual use, shrink to one or two points. Including too few steps loses differentiating information because people who use the same response category may differ if intermediate steps were included." Therefore, the scaling of the new instrument assessing teacher efficacy in measurement and evaluation practices was in between (i.e., neither a 100-point format nor 5-point Likert scale) as being 9-point.

3.5.2.2. Pilot Study

The pilot study was conducted by administrating the instrument to 118 elementary and secondary school teachers in Ankara. Twenty-three percent (23%) of these teachers were teaching at elementary level, while 77% was working in secondary level. There were 65 female teachers and 53 male teachers. The average age and teaching experience in years was 40 and 16, respectively. Nearly, half of the

sample (49.2%) was composed of graduates of faculties of education. Approximately 24% of the participant teachers had a science (e.g. physics, biology, and chemistry) or mathematics major, whereas 76% of them were teaching social sciences (e.g., teaching history, languages like Turkish or English, or classroom teacher). Among all participants, 12% of them have taken a course on measurement and evaluation during their university education and 68% of them have participated an in-service training on measurement and evaluation.

To examine the factor structure of TEMES, Exploratory Factor Analysis (EFA) was performed through SPSS 15.0. Before the analysis, the researchers checked the assumptions of Exploratory Factor Analysis, which were proof of metric variables, correlations above .30, Bartlett's Test of Sphericity, KMO (Kaiser-Mayer Olkin) value (>.60), multivariate normality, and absence of outliers (Hair, Anderson, Tatham, & Black, 2006). The instrument was a 9-point scale and the responses were regarded as efficacy scores (metric variable) for each participant. There was no correlation coefficient which was less than .30. Bartlett Test resulted in a significant value which meant that correlation matrix was significantly different than an identity matrix, i.e., none of the correlations between the items were zero (Tabachnick & Fidell, 2007). Moreover, KMO value (.93) was exceeding the criterion value of .60. Before examining multivariate normality, univariate normality was checked by observing skewness and kurtosis values, significance of Kolmogorov-Smirnov and Shapiro-Wilk Tests and histograms with normal curves. The skewness and kurtosis values were between +3 and -3 (Tabachnick & Fidell, 2007)**,** but Kolmogorov-Smirnov and Shapiro-Wilk Tests were significant which indicated that distribution differed from normality. Yet Kolmogorov-Smirnov and Shapiro-Wilk Tests are conservative tests, the researchers continued to examine univariate normality by checking histograms and they noticed that univariate normality was not violated according to the histograms with normal curves.

In addition to univariate normality, existence of multivariate normality was tested by running norm test macro in SPSS 15.0. This analysis yielded Small Test with a significant result showing the violation of multivariate normality but this test was a kind of Chi-Square Test and it was sensitive to sample size. Cases which have Mahalonobis Distance values larger than the critical value (45.51 for $\alpha = .05$ and $df = 24$) were checked to detect multivariate outliers. Only three out of 118 cases were extreme cases. Boxplots were also examined to determine whether there was any univariate outlier. It was seen that there were no serious outlier in any of the cases. These results showed that it is possible to continue factor analysis.

Factor analysis resulted in that the new instrument had two factors which were named as: *developing measurement and evaluation tools* and *applying and analyzing the results of measurement and evaluation tools.* Approximately 62% of the variance in teachers' efficacy toward measurement and evaluation tools was explained by these two factors. The scree plot, also suggesting two factors, is presented in Figure 3.1. Based on the finding of the pilot study, none of the items were eliminated. To report on reliability of the two factors, Cronbach Alpha Coefficients were calculated and resulted in following values .95 and .93, respectively.

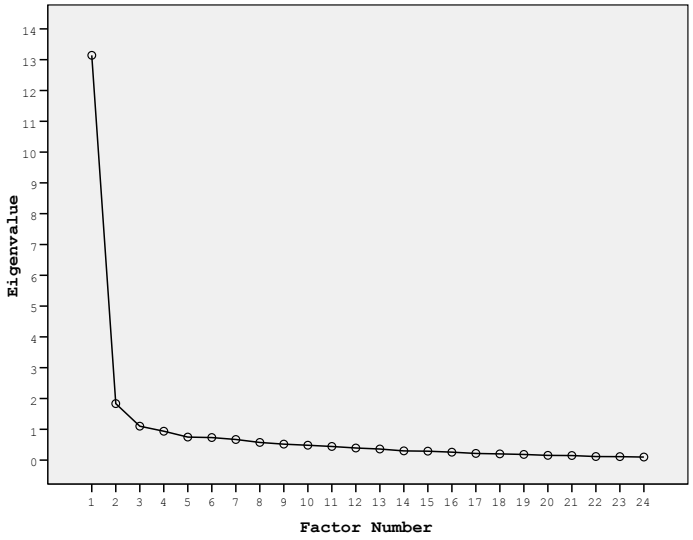

Figure 3.1 *Scree Plot*

Items loaded on the related factors with high values and this indicated exploratory factor analysis was appropriate for the instrument. Some items of the first factor with the factor loadings were as follows: “How well can you develop appropriate questions for instructional content?” (-.95) and “How well can you gauge student comprehension of what you have taught?” (-.80). Factor loadings found for some of the items of the second factor were: “How well can you prepare individual measurement and evaluation activities (e.g. performance evaluation, project)?” (.81) and “How well can you develop alternative measurement and evaluation tools (e.g., concept maps, constructed grid)?” (.87).

Reliability analysis for Alternative-ME and Traditional-ME yielded following coefficients: .89 for *Alternative-ME* and .69 for *Traditional-ME*. Item total correlations ranged from .41 to .76 for Alternative-ME and from .34 to .68 for Traditional-ME, indicating that all the items were working as intended.

3.5.3. Scale for Measuring Frequency of Using Different Measurement and Evaluation Tools

The purpose of developing a scale including all measurement and evaluation tools was to measure how frequently teachers use different measurement and evaluation tools. Herewith the researchers constructed an instrument which was a 5-point Likert scale (ranging from never to always) including 17 measurement and evaluation tools that were proposed by the Turkish Ministry of National Education (MoNE) in the latest curriculum (Erdoğan, 2007). Tools were classified as traditional and alternative measurement and evaluation in this scale. In order to see the difference between using alternative and traditional measurement and evaluation methods in terms of the effect of teachers' sense of efficacy toward measurement and evaluation practices, the researchers derived two variables from this scale as Traditional-ME (mean score of the items including traditional assessment methods) and Alternative-ME (mean score of the items including alternative assessment methods). Teachers were asked to indicate their frequency of using listed measurement and evaluation tools out of five frequency choices as *never*, *rarely*, *sometimes*, *frequently*, and *always*. Ten items measuring Alternative-ME asked for the frequency of using word matching, written reports, interview with students and observation, drama, portfolio, concept map, constructed grid, performance evaluation, self-report, and peer evaluation. Traditional-ME was measured by seven items asking how frequently teachers used open-ended questions, short answered questions, multiple choice test, true/ false questions, matching questions, fill in type, and question-answer technique.

In the pilot study, EFA was conducted to define whether items measuring frequency of using alternative tools could be differentiated from the ones measuring the frequency of using traditional tools. EFA findings indicated that this scale had two factors as having expected items relating to alternative and traditional separately. Reliability analysis revealed the following coefficients for

frequency of using alternative and *frequency of using traditional tools* respectively: .89 and .69.

3.5.4. Turkish Teachers' Sense of Efficacy Scale

The instrument (previously called as *Ohio State Teacher Efficacy Scale,* now known as *Teachers' Sense of Efficacy Scale* which was developed by Tschannen-Moran and Woolfolk Hoy (2001), included three factors: efficacy for student engagement, efficacy for instructional strategies, and efficacy for classroom management. Tschannen-Moran and Woolfolk Hoy (2001) examined their scale in three studies with different pre-service and in-service teachers (the sample sizes were 224, 217, and 410, respectively). In consideration with the factor loadings, some items were extracted from the scale and the researchers decided to continue with 32-item scale after the first study. In the second study, number of items decreased to 18 and factor analysis resulted in a 3-factor structure, and the number of items in each factor was as follows: 8 items in *efficacy student engagement* (ESE), 7 items in *efficacy for instructional strategies* (EIS), and 3 items in *efficacy for classroom management* (ECM). Tschannen-Moran and Woolfolk Hoy (2001) designed one more study with 410 participants to refine *Teachers' Sense of Efficacy Scale.* The final reported reliability coefficients for the 3-factor scale were as follows: .81 for ESE, .86 for EIS and .86 for ECM. Each factor has 8 items. Çapa, Çakıroğlu, and Sarıkaya (2005) adapted *Teachers' Sense of Efficacy Scale* in Turkish by administering the translated version to 628 pre-service teachers in six faculties of education in Turkey. Çapa and her colleagues (2005) found that the adapted version of TSES was also composed of three factors as ESE, EIS, and ECM with similar reliability estimates ranging from .82 to .86.

3.6. Data Collection Procedure

After the scale was developed, necessary documents were submitted to the METU Human Subjects Ethics Committee (HSEC). While waiting for the decision of the committee, the researchers made a random list of schools from the complete school list of the Turkish Ministry of National Education. The study was conducted in three different cities: Ankara (the districts of Çankaya and Sincan), Samsun (Center district), İstanbul (the districts of Eyüp, Bakırköy and Zeytinburnu). The instrument and proposal were submitted to Educational Research and Development Head Office (ERDHO) after the METU HSEC approved that the study has an applicable instrument and there is no problem with the design for the ethical considerations.

Questionnaires were prepared in an optic format to make both data collection and entry process easier and quicker. Listed schools in Ankara, Samsun, and İstanbul were visited by the researchers and questionnaires were filled by the teachers. During the data collection process, the researchers observed the participants to see whether they responded the instrument independently and the researchers answered the questions of the participants to prevent missing data. Data collection lasted 10 to 15 minutes.

3.7. Data Analysis

The following points suggested by Meyers, Gamst, and Guarino (2006) were considered before the data analysis: Is there any missing or incorrect data entry? Is there a pattern for missing data? Are there any extreme values that may affect the results of the study? Are the assumptions of the intended multivariate statistical techniques met? What can be done if any of these assumptions is violated somehow?

First of all, data were screened to check for missing values and for incorrect data entry if any existed. No incorrect entry was detected, but both in demographic variables and scale items, there were some missing values not exceeding 5 percent. Moreover, it was found that missingness followed a random pattern by running *Little's MCAR Test* (Little & Rubin, 1987). Therefore, researchers decided to impute the missing values by using the Expectation Maximization (EM) algorithm (Tabachnick & Fidell, 2007). Tabachnick and Fidell (2007) reported that this method is a commonly used one when missing values are at random. In Expectation Maximization, two steps are followed: estimation of missing values and then estimation of parameters by regression analysis (Hair et al., 2006). In addition, Allison (2002) reported that EM was practical because it checked for all appropriate variables to impute missing values.

Second, after missing value analysis was completed, unique scores were extracted for each scale, i.e., *Teacher Self-Efficacy toward Measurement and Evaluation Practices Scale (TEMES), Frequency of Using Different Measurement and Evaluation Tools Scale (FMES)* and *Turkish Teachers' Sense of Efficacy Scale (TTSES).* Four mean scores were calculated for the participants: SE-Mean for self-efficacy toward measurement and evaluation practices, Alternative-ME for frequency of using alternative measurement and evaluation tools, Traditional-ME for frequency of using traditional measurement and evaluation tools and TTSES-Mean for teacher efficacy.

Third, data were collected from teachers who were teaching at elementary and secondary schools in Ankara, Samsun, and Istanbul. Therefore, whether teachers' responses differed in consideration with the city difference was examined by conducting One-way Analysis of Variances (One-way ANOVA) before further analyses. In this study, the researchers set the level of significance (α) at .05.

Fourth, to provide validation evidence for TEMES, Confirmatory Factor Analysis (CFA) was conducted by Analysis Moments of Structures (AMOS) 4.1. CFA has a deductive approach in that the aim is to find out the factorial structure which theoretical framework supports (Meyers et al., 2006). Bollen and Long (1993) summarized the steps of CFA which were model specification, model identification, model estimation, model evaluation, and model respecification, respectively. In the first step, model specification, researchers develop a model in consideration with the theory, and then check for whether the model can be identified in the model identification step. Model identification compares the number of variables in the analysis and the number of parameters estimated by the model. The difference between these two is known as *degrees of freedom* (df) and this value should be positive to indicate that the model is identified. In the third step, model estimation, the specified (theoretical) model is compared to what the data represents (observed model) by the statistical program, AMOS 4.1 in this research study. Then, model evaluation includes deciding about whether model fits the data by evaluating what the analysis yields, i.e., fit indices (e.g. NNNFI, CFI, and RMSEA), chi-square goodness-fit test results, unstandardized and standardized parameter estimates. According to these values, researchers can change or maintain the estimated model. When they add or delete some connections in the model, this is named as model respecification.

Next, Cronbach's Coefficient Alpha was computed to check for the internal consistency of TEMES, TTSES, Alternative-ME, and Traditional-ME. Estimated scale reliabilities in the case of any item deleted were also examined to check whether there is any problem with the items.

Regarding the examination of whether TEMES is an appropriate instrument to measure teacher self-efficacy toward measurement and evaluation practices, Canonical Correlational Analysis was conducted the to examine the relationship between the factor scores of TTSES (Turkish Teachers' Sense of Efficacy Scale)

and the factor scores of TEMES (Teacher Self-Efficacy toward Measurement and Evaluation Practices Scale). In canonical correlation analysis, correlations between variables in and between the two sets are examined to understand the relationship in and between the sets. In each set, variables are loaded on a related canonical variate and canonical correlations above .30 are the concern of a researcher. Then, to examine the effect of gender and teaching level on the factors of Teacher Self-Efficacy toward Measurement and Evaluation Practices Scale (TEMES), Multivariate Analysis of Variance (MANOVA) was run. These analyses were performed using SPSS 15.0.

Finally, to find out answers for the research questions, Structural Equation Modeling (SEM) was conducted by AMOS 4.1. The structural model was specified according to the theoretical framework which is derived from the related literature on teachers' sense of efficacy. The corresponding variables were year in teaching and frequency of using alternative and traditional measurement and evaluation tools in this study. The model is represented in Figure 3.2. SEM is advantageous in terms of assessing and controlling measurement errors (Meyers et al., 2006). In this analysis, there are mainly two models named as *structural* and *measurement*. While measurement model specifies the relationship between the latent (unobserved) and manifest (observed) variables, the structural model identifies the relationship among the latent variables (Byrne, 2001). SEM uses maximum likelihood method which estimates the values of parameters that would provide the maximum likelihood of observed data to the theoretical model. In SEM analysis, comparison is made between the theoretical model and the model which is presented by the observed data. This comparison is carried out by examining the fit indices, chi-square test, and correlational estimates to conclude whether the theoretical model fits the collected data (Meyers et al., 2006). In this study, the researchers checked chi-square statistic (Hoyle, 1995) and root mean square error of approximation (RMSEA; Steiger & Lind, 1980) known as absolute

fit indices; in addition to the comparative fit index (CFI; Bentler, 1990) and non-normed fit index (NNFI; known as Tucker-Lewis Index; Bentler & Bonett, 1980) which were categorized as incremental fit indices (Hair et al. 2006). For both of the absolute and incremental fit indices, there are some criteria to evaluate the model fit. If Chi-square statistic results in significant value, then the specified model is different than observed data; that is, the model does not fit the data. However, chi-square measure is sample size dependent. Therefore, it is better to check for other fit indices to understand the model fit (Hair et al., 2006). Browne and Cudeck (1993) reported that close fit is indicated by RMSEA values lower than .05; mediocre fit is indicated by the values between .05 and .08; and poor fit is indicated by the values over .10. Later, the criteria of mediocre fit and poor fit for RMSEA were defined as values between .08 and .10 is an evidence for mediocre fit and the values higher than .10 indicate poor fit (MacCallum, Browne, & Sugawara, 1996). In addition to these criteria for absolute fit indices, CFI and NNFI changes between 0 and 1 (Hair et al., 2006), and should be greater than .95 to indicate good fitting model (Hu & Bentler, 1999).

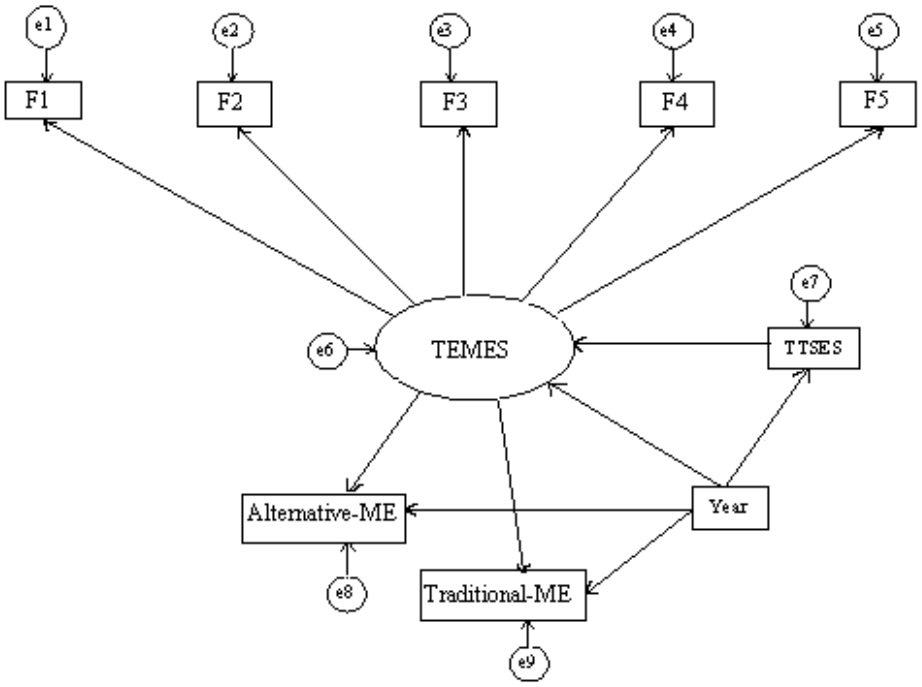

Figure 3.2 *Structural Model Displaying the Relationship between Variables*

3.8. Limitations

The following limitations are associated with this study:

1. Correlational research was used in this study; therefore, no causal relationship can be made between the research variables.
2. The present study is limited with the relationship between year in teaching, frequency of using different measurement and evaluation tools, and teacher self-efficacy toward measurement and evaluation tools. There may be other variables related to teacher self-efficacy toward measurement and evaluation practices.

3. The present study is relied on self-report data. Resources such as observation reports, interview reports, or peer evaluation are not used, because of the quantitative nature of the study.
4. The present study is limited with the teachers who have the characteristics which are defined in section 3.4. Data were collected from the teachers who work in public primary and high schools in Çankaya and Sincan districts of Ankara, city center of Samsun, Eyüp, Zeytinburnu, and Bakırköy districts of İstanbul. Due to convenience sampling is preferred, the results does not represent all the teachers in Turkey.

CHAPTER IV

RESULTS

In this chapter, results of data analysis are presented under the following headings: Descriptive statistics of scale scores (for *TEMES, FMES* and *TTSES* scales), examination of related assumptions for further analyses, the results of one way ANOVAs, and separate Confirmatory Factor Analysis for TEMES and TTSES, reliability coefficients, additional validity evidences including results of MANOVA and Canonical Correlation Analysis, and results of Structural Equation Modeling.

In this study, the purpose was to explore the relationship between teachers' sense of efficacy beliefs toward measurement and evaluation practices, teachers' sense of efficacy, year in teaching, and frequency of using different measurement and evaluation tools.

Before the further analyses, the researchers supposed that it was practical to examine whether the items differed significantly when city was considered as an independent variable by conducting one way ANOVA for each item of three scales. This is performed because one-way ANOVAs provided the researchers an opportunity to evaluate mean differences between the data of three cities. To make sure that the data were appropriate for running separate one-way ANOVAs, the researchers checked for the corresponding assumptions which were independent observation, normality, and equality in population variances (i.e., homogeneity of variances) (Gravetter & Wallnou, 2007). The researchers prevented participants' responses not to be affected each other by being present at where the data were

collected; therefore, the assumption of independent observation is verified. To check that normality assumption was met, skewness and kurtosis values for each item of three scales, and histograms with normality curves were examined. The researchers concluded that there was no problem with the normality assumption; hence, there were only two items which had kurtosis values exceeding the criteria of being in between -3 and +3 (Tabachnick & Fidell, 2007) and normality curves indicated no skewed distributions. Moreover, Levene's Test yielded nonsignificant value and this indicated that there was no difference between error variances across the data of different cities. After preliminary analysis, one way ANOVAs for each item was run separately, and results indicated that only three of the 65 items differed significantly, but effect sizes were pretty low (ranging from .02 to .03). Therefore, data of three cities were gathered and totally 394 cases were analyzed in this study.

Mean, standard deviation, minimum and maximum values for the study scales, TEMES, TTSES, Alternative-ME, and Traditional-ME were computed and displayed in Table 4.1.

Table 4.1

Results of Descriptive Statistics for TEMES, TTSES, and FMES

Variables	*M*	*SD*	*Min*	*Max*
TEMES	6.83	.98	1	9
TTSES	6.96	.82	1	9
Alternative-ME	2.85	.84	1	5
Traditional-ME	3.48	.69	1	5

Descriptive statistics indicated that the mean scores of teachers' sense of efficacy (M_{TTSES}=6.96) which was assessed by factors of efficacy in student engagement, instructional strategies, and classroom management), and teacher self-efficacy toward measurement and evaluation practices *(*M_{TEMES}=6.83) are approximate to each other. TEMES (*SD*=.98) and TTSES (*SD*=.82) scores have approximately the same variation. Moreover, traditional (*M*=3.48) and alternative (*M*=2.85) measurement and evaluation tools are used in different frequency by the teachers who work in public elementary and secondary schools. Traditional measurement and evaluation tools were reported to be used more frequently than alternative ones. The variation of Alternative-ME scores (*SD*=.84) is slightly higher than the variation of Traditional-ME scores (*SD*=.69).

4.1. Confirmatory Factor Analysis

Researchers proposed a five-factor structure for TEMES based on the levels of measurement and evaluation practices. These factors were determining assessment goals and techniques, developing measurement and evaluation tools, administering measurement and evaluation tools and evaluating the results, analyzing the results, and using and sharing results in other courses.

CFA resulted in significant chi-square value (=221.42), CFI value of .99, and NNFI value of .97; but RMSEA value was close to .10 (=.095) and this indicated poor fit (MacCallum, Browne, & Sugawara, 1996). Therefore, researchers checked the modification indices (i.e., error covariance) of errors, and detected the ones with high values, i.e., most striking values among all (Arbuckle, 1999). The pairs with high error covariances were ε_6- ε_{22}, ε_8- ε_{13}, ε_9- ε_{10}, ε_9- ε_{15}, ε_{15}- ε_{16}, and ε_{20}- ε_{21}. The items related to these errors were examined in terms of belonging to the same factor or measuring related tasks in measurement and evaluation practices. The following item pairs loaded on same factors, namely item 8 and 13 loaded on the second factor, item 9 and item 15, item 15 and item 16 loaded on

the fourth factor, and item 20 and item 21 loaded on the fifth factor. Although two of the item pairs, 9-10 and 6-22, did not load on the same factors, these items measured similar or consequent tasks in measurement and evaluation practices. For example, both item 6 and item 22 asked for determining and developing alternative measurement and evaluation tools. In that sense, related error pairs were connected in the model and analysis was run again. After this change, RMSEA value decreased to .08 and this value indicated mediocre fit (MacCallum, Browne, & Sugawara 1996). In addition, resulting NNFI (.98) and CFI (.98) values supported good fitting model due to being higher than .95 (Hu & Bentler, 1999). Moreover, chi-square statistics resulted in a significant value of 870.60 ($p<.00$). Although this indicated that the CFA model unlikely representing a good fit, the researchers considered the result which is proved by other fit indices, CFI, NNFI, and RMSEA. Because chi-square statistic is sensitive to sample size and other fit indices are took into consideration in the case of significant chi-square result (Byrne, 2001). Figure 4.1 represents the final CFA model with standardized estimates ranged from .57 to .77.

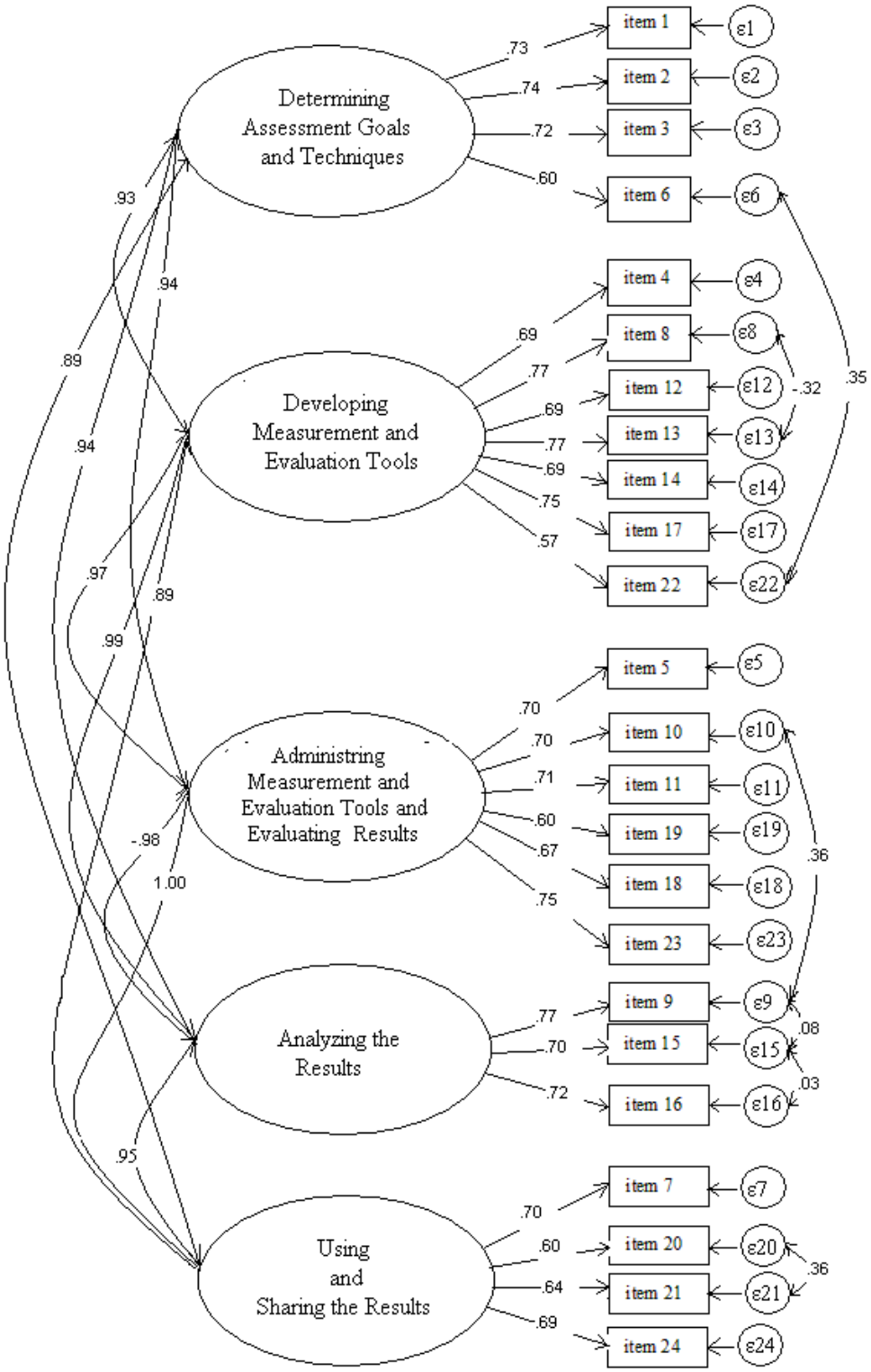

Figure 4.1 *Five Factor CFA Model of TEMES with Standardized Estimates*

4.2. Reliability

Cronbach Alpha Reliability Coefficient of each factor was computed. Table 4.2 displays the reliability coefficients along with the reliability coefficients if item deleted. The reliability coefficients for each factor of TEMES were as follows: .76 for *determining assessment goals and techniques,* .87 for *developing measurement and evaluation tools,* .85 for *administering measurement and evaluation tools and evaluating the results,* .80 for *analyzing the results,* and .80 for *using and sharing results*.

When alpha if item deleted column was examined, it appeared that most of the items were contributing to the corresponding factor. Only two of the items seemed problematic. If item 6 (e.g., "How well can you determine the alternative measurement tools for multifaceted evaluation?") is deleted, the alpha value will increase to .80 from .76. In addition, in case of deleting item 9, which is questioning "How well a teacher can achieve to test the validity and reliability of a measurement and evaluation tool" alpha value for the fourth factor will increase from .80 to .83. On the other hand, neither in the first factor nor the fourth one the increase in the reliability coefficient was too much. Besides, the researchers and experts, who examined the items in terms of content, agreed on keeping both the item 6 and the item 9 in the scale as it is. Therefore, the researchers decided not to delete or change any item according to the change in factor reliabilities in the case of item deletion.

Table 4.2

Reliability Coefficients of TEMES Factors and Related Items

	Reliability	*Alpha If Item Deleted*
Determining assessment goals and techniques	.76	
Item 1		.67
Item 2		.65
Item 3		.70
Item 6		**.80**
Developing measurement and evaluation tools	.87	
Item 4		.85
Item 8		.84
Item 12		.85
Item 13		.85
Item 14		.85
Item 17		.85
Item 22		.87
Administering measurement and evaluation tools and evaluating results	.85	
Item 5		.83
Item 10		.81
Item 11		.82
Item 18		.84
Item 19		.82
Item 23		.82
Analyzing the results	.80	
Item 9		**.83**
Item 15		.69
Item 16		.66
Using and sharing the results	.80	
Item 7		.78
Item 20		.71
Item 21		.73
Item 24		.76

In addition to the reliability analysis of TEMES, researchers examined the reliability coefficients of the factors of FMES. They realized that the reliability coefficients of these factors were not quite different than those yielded in the pilot study. The reliability coefficient of Alternative-ME was decreased from .89 to .86 and that of Traditional-ME was increased from .69 to .70. All of the items were contributing to the corresponding factors of FMES.

Lastly, the reliability analysis of the TTSES was conducted. The reliability coefficient was found as .93, when the TTSES was considered as one-dimensional. In the adaptation study of the TTSES, the reliability of efficacy scores was also found as .93 (Çapa, Çakıroğlu, and Sarıkaya, 2005).

4.3. Additional Validity Evidence

After checking the factor structure of *Teacher Self-Efficacy toward Measurement and Evaluation Practices Scale* (TEMES), researchers ran Canonical Correlational Analysis between the factor scores of TTSES (Turkish Teachers' Sense of Efficacy Scale) and the factor scores of TEMES in order to represent as additional validity evidence. The rationale behind running canonical correlation between the factors of two scales was that TTSES is a valid and reliable measure to assess teacher self-efficacy in student engagement, instructional strategies, and classroom management and high correlation coefficients between the factors of TEMES and TTSES would indicate that TEMES is a valid and reliable measure to assess teachers' sense of efficacy as well. But before running canonical correlation analysis, Confirmatory Factor Analysis (CFA) was performed to check whether TTSES resulted in three factors as in original version.

Moreover, two separate Multivariate Analysis of Variance (MANOVAs) were conducted to examine the factors of TEMES in terms of gender and teaching level. These analyses were also done in order to provide support for validity.

4.3.1. Canonical Correlation between TTSES and TEMES

Before examining the relationship between TEMES and TTSES through canonical correlation, CFA was performed to check whether TTSES resulted in three factors as in original version. There were three factors named as *efficacy toward classroom management, efficacy toward instructional strategies, and efficacy toward student engagement.* CFA findings indicated that three-factor model fits the data. For the confirmatory model represented in Figure 4.2, standardized estimates ranged from .52 to .79. Fit indices were resulted as follows: .08 for RMSEA, .83 for CFI, and .79 for NNFI. These values indicated mediocre fit of the confirmatory model to the TTSES data. Moreover, the items of TTSES were found to be significantly loaded on the corresponding three factors.

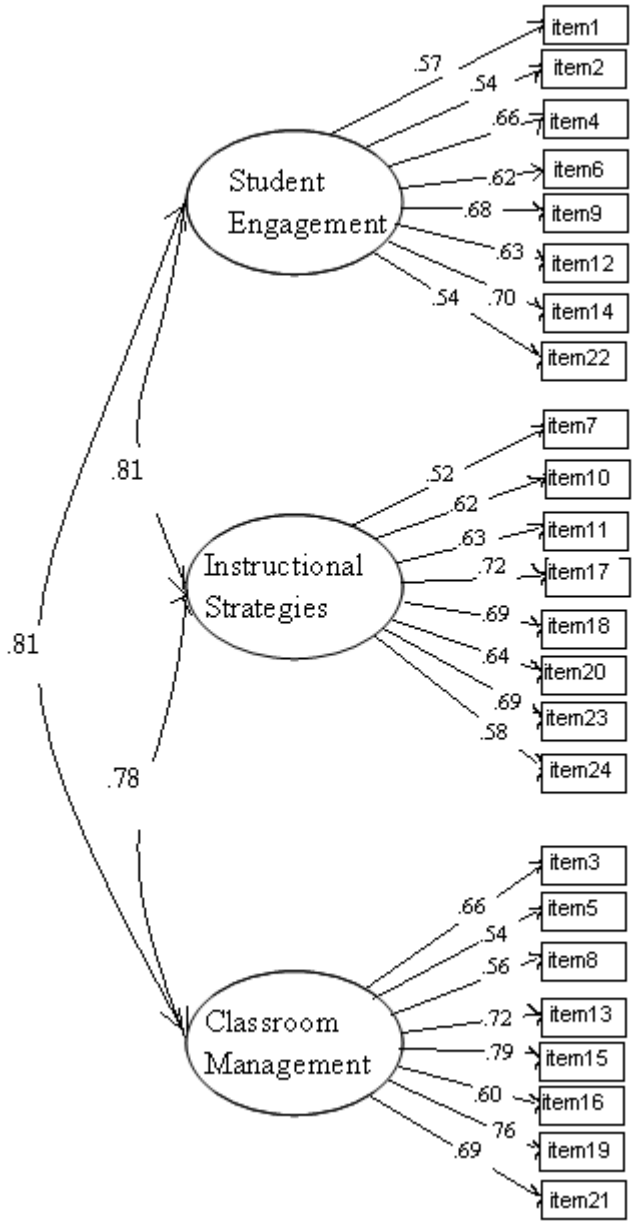

Figure 4.2 *Factor Structure of TTSES with standardized estimates*

Next, Canonical Correlation Analysis was run between the factor scores of TTSES (i.e., score for efficacy in student engagement, classroom management and instructional strategies) and TEMES (i.e., score for efficacy in determining

assessment goals and techniques, developing assessment tools, administering assessment tools and evaluating the results, analyzing the results and using/ sharing the results). TTSES was considered as the first covariate with the factor scores represented by F_1, F_2 and F_3 and TEMES was the second covariate with the factor scores named as f_1, f_2, f_3, f_4, and f_5 in the Canonical Correlation Representation (Figure 4.3).

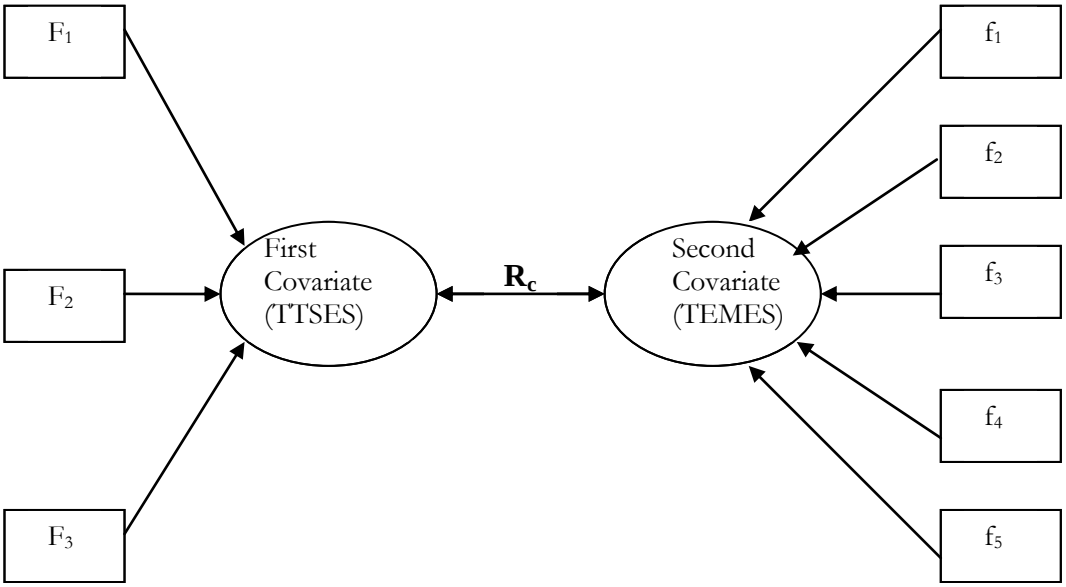

Figure 4.3 *Canonical Correlation Representation between the Factors of TTSES and TEMES*

In this study, researchers found that there was no multicollinearity between factors of TTSES and factors of TEMES, because the correlation coefficients did not exceed .90. In that sense, the researchers agreed that canonical correlation analysis is appropriate to examine the relationship between the factors of two scales.

Canonical Correlation Analysis revealed a canonical correlation coefficient (R_c= .63) which was higher than .30 (Tabachnick & Fidell, 2007) for the first canonical

pair (Table 4.3). This result indicated that there was a high and positive correlation between the factor scores of both scales. Furthermore, canonical loadings for the factors of TTSES and TEMES were higher than .30 and had the same sign which proved that change in teachers' sense of efficacy in student engagement, classroom management, and instructional strategies was significantly parallel to the change in efficacy in determining assessment goals and techniques, developing assessment tools, administering assessment tools and evaluating the results, analyzing the results and using/ sharing the results. For example, teachers who are efficacious in student engagement tend to have high self-efficacy in assessing student learning and evaluating the results, vice versa.

Table 4.3

Results for Canonical Correlation Analysis between the Factors of TTSES and TEMES

	First Canonical Variate	
	Correlation	Coefficient
TTSES		
F_1	-.76	-.13
F_2	-.99	-.85
F_3	-.77	-.07
Percent of Variance	.72	
Redundancy	.28	
TEMES		
f_1	-.93	-.42
f_2	-.87	.04
f_3	-.94	-.39
f_4	-.87	-.19
f_5	-.84	-.14
Percent of Variance	.79	
Redundancy	.31	
Canonical Correlation	.63	

4.3.2. Multivariate Analysis of Variance: Investigation of TEMES by Gender and Teaching Level

The researchers examined teacher self-efficacy toward measurement and evaluation practices by gender and teaching level through Multivariate Analysis of Variance (MANOVA). Two separate MANOVAs were run for each independent variable, because no correlation was found between the independent variables. First MANOVA was run for the independent variable of gender, and then for the independent variable of teaching level. The dependent variables were five factors of TEMES; efficacy for determining assessment goals and techniques, efficacy for developing assessment tools, efficacy for administering assessment tools and evaluating the results, efficacy for analyzing the results, and efficacy for using/ sharing the results.

TEMES Factors according to gender difference. In this analysis, gender was the independent variable and five factors of TEMES were the dependent variables. Before the running MANOVA, related assumptions, i.e., independent observation, univariate and multivariate normality and outliers, homogeneity of population covariance matrix for dependent variables, and metric dependent variables (Tabachnick & Fidell, 2007) were examined. Independent observation was met by researchers' attendance during data collection. To check the normal distribution of TEMES factors, the researchers examined skewness and kurtosis values, Kolmogorov-Smirnov and Shapiro-Wilk Tests, and histograms with normality curves. Skewness and kurtosis values ranged from -.71 to 1.01, and this was an evidence for normal distribution because normality requires skewness and kurtosis values to be in the range of -3 to 3 (Tabachnick & Fidell, 2007). In addition, Kolmogorov-Smirnov and Shapiro-Wilk Tests resulted in significant values and significant results of these tests indicate non normality. On the other hand, as these tests are conservative (Field, 2006), the researchers checked the histograms with normal curves in addition to skewness and kurtosis as evidence of

normal distribution. Histograms of each TEMES factor displayed a slight skewed distribution; nevertheless, the researchers concluded that univariate normality is met. Multivariate normality was checked by running norm test macro in SPSS, and it ensured the researchers examining Mardia's test, plot of ordered square distances, and Mahalonobis Distances to check multivariate outliers. Mardia's test yielded significant result and that means the distribution of TEMES factors' mean scores distributed non-normally (Tabachnick & Fidell, 2007). In addition, the plot of ordered square distances and the cases exceeding the critical value of χ^2 (5) = 24.57 were examined, it was found that there were only five cases. These cases were not deleted because decrease in sample size limits the generalizability (Tabachnick & Fidell, 2007). Moreover, these cases were reviewed and no problem was found in terms of data entry or demographic characteristics. Then, homogeneity of covariance matrices were checked by Box's M and Levene's Tests. Box's M resulted in a significant value (30.37, $p<.01$) and this indicated that observed covariance matrices of TEMES factors were unequal across groups (Field, 2006). However, Box's M is sensitive to non-normality. Therefore, Levene's Test was examined and it yielded non-significant result for each TEMES factor. This result was an evidence for homoscedasticity or equal observed covariance matrices of TEMES factors. Lastly, whether dependent variables were metric or continuous was considered. For this multivariate analysis, the dependent variables were mean scores of TEMES's five factors. These variables were already metric; hence, the researchers computed the factor scores by taking average of all participants' scores for each factor.

After related assumptions check were completed, descriptive statistics were run to describe basic characteristics of participating teachers. Descriptive statistics indicated that the mean score of male teachers was lower than that of female teachers except for the fourth factor. In terms of efficacy in analyzing the results of measurement and evaluation, male teachers reported higher self-efficacy

compared to female teachers. In conclusion, the researchers estimated that the difference between the mean scores of each factor was not excessive in terms of gender difference. The descriptive results are displayed in Table 4.4. Unequality in cell sizes was detected when the descriptive statistics output was examined in detail. Type III Adjustment is a way of solving the problem of unequal sample size in each cell. It is applicable to both balanced and unbalanced designs, and it works not considering whether there is an interaction between independent variables or not (Tabachnick & Fidell, 2007). Therefore, the researchers ran Type III Adjustment to overcome the unequal cell sizes.

Table 4.4

Results of Descriptive Statistics

		F_1		F_2		F_3		F_4		F_5	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Female	225	6.87	1.06	6.67	1.11	6.91	1.04	6.93	1.20	7.01	1.10
Male	169	6.78	.96	6.60	1.10	6.90	1.11	7.00	1.13	6.81	1.11

Findings of MANOVA are presented in Table 4.5. Wilk's lambda is the most commonly preferred multivariate test statistic (Hair, Anderson, Tatham, & Black, 2006); however, the researchers examined the result of Pillai's Trace Test. This is because Pillai's Trace Test is preferred to Wilk's Lambda in case of the heterogeneity of covariance matrices. MANOVA resulted in a significant Pillai's Trace Test with $F(5, 388) = 2.41$, $p < .05$, $\eta^2 = .03$.

Table 4.5

MANOVA for TEMES Factors by Gender

		ANOVA				
	MANOVA	F_1	F_2	F_3	F_4	F_5
	F(5,388)	F(1,392)	F(1,392)	F(1,392)	F(1,392)	F(1,392)
Gender	2.41*	.64	.45	.01	.31	3.00

* *p<.05*

After multivariate analysis, univariate statistics were examined. In order to correct for the increase in the overall Type I error rate. Bonferronni correction was used. Bonferronni correction is a kind of method controlling family wise error rates by dividing the alpha by the number of comparisons (Field, 2006). In this case, Bonferronni correction was administered by dividing the alpha by five (equal to the number of dependent variables) and the result of univariate analysis (ANOVA) was checked according to this new alpha value (α= .01). The univariate analysis resulted in nonsignificant difference between factors of TEMES in consideration with gender, and following F values were found for each factor: F_{F1} (1,392) = .73, η^2= .00, F_{F2} (1,392) = .44, η^2= .00, F_{F3} (1,392) = .01, η^2= .00, F_{F4} (1,392) = .36, η^2= .00, and F_{F5} (1,392) = 3.03, η^2= .01. This was an expected result, because Çakan (2004) found out that teachers' perception level of qualifications in assessment did not differ by gender and the descriptive statistics had already pointed out that factor scores of TEMES were approximately similar for female and male teachers.

TEMES Factors according to the difference in teaching level. In the second MANOVA, teaching level was the independent variable and five factors of TEMES were the dependent variables. The researchers previously examined and

reported the related assumptions of MANOVA, i.e., independent observation, univariate and multivariate normality and outliers, homogeneity of population covariance matrix for dependent variables, and metric dependent variables (Tabachnick & Fidell, 2007) for the same dependent variables. In addition, homogeneity of covariance matrices was checked and it was found that Levene's Test was nonsignificant for each factor. Researchers retained the null hypothesis of this test; that is, error variances of dependent variables were equal across groups (Tabachnick & Fidell, 2007).

The researchers agreed on that no assumption of MANOVA was violated, and continued evaluating the results of analyses with descriptive statistics. Descriptive statistics for TEMES factors in terms of difference in teaching level summarized in Table 4.6. Secondary school teachers were more efficacious than elementary school teachers in determining assessment goals and techniques (first factor), developing measurement and evaluation tools (second factor), and analyzing the results (fourth factor). However, there was no difference between the teachers working at different teaching levels in terms of efficacy in administering measurement and evaluation tools and evaluating the results. In addition, elementary school teachers were more efficacious than secondary school teachers toward using and sharing results of classroom measurement. The researchers noticed that the change in mean score of the fourth factor between secondary school teachers (M= 7.14, SD= 1.02) and elementary school teachers (M=6.80, SD= 1.26) was greater compared to change in the mean scores of other factors. As in previous MANOVA, unequality in cell sizes was detected in the output and Type III Adjustment was selected while conducting the main analysis.

Table 4.6

Results of Descriptive Statistics

		F_1		F_2		F_3		F_4		F_5	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Elementary	210	6.77	1.11	6.62	1.13	6.91	1.14	6.80	1.26	6.94	1.10
Secondary	184	6.91	.92	6.65	1.04	6.91	.99	7.14	1.02	6.90	1.11

Table 4.7 illustrates F-statistics for both multivariate and univariate analysis. Multivariate statistics resulted in Pillai's Trace Value of $F(5, 388) = 5.53$, $p<.05$, $\eta^2=.07$. By administering Bonferronni correction, the level of alpha was set to .01. The results of univariate analysis indicated that significant difference was only found for the fourth factor, $F_{F4}(1,392) = 8.00$, $p<.01$, $\eta^2=.02$. In order to determine at which teaching level participating teachers scored significantly more, the contrast table in MANOVA output was examined, and it was noticed that secondary school teachers were more efficacious in analyzing the results of measurement and evaluation compared to elementary school teachers with mean difference of .33. Though the effect size was not too high, the researchers considered the significance of the effect of teaching level on teacher self-efficacy in analyzing the results of measurement and evaluation, because this result was confirmed in the literature by the finding of Çakan (2004) that secondary school teachers' perception level of qualifications in assessment was higher than elementary school teachers.

Table 4.7

MANOVA for TEMES Factors by Teaching Level

		ANOVA				
	MANOVA	F_1	F_2	F_3	F_4	F_5
	F (5,388)	F (1,392)	F (1,392)	F (1,392)	F (1,392)	F (1,392)
Teaching Level	5.53*	2.12	.07	.00	8.00**	.12

* *p<.05, ** p<.01*

4.4. Structural Equation Modeling

In this study, the researchers' purpose was to answer the following research problems: Is year in teaching experience a significant predictor for frequency of using alternative and traditional measurement and evaluation tools? Do year in teaching experience and teachers' sense of efficacy significantly predict teacher self-efficacy in measurement and evaluation practices? Does teacher self-efficacy in measurement and evaluation practices predict the frequency of using alternative and traditional measurement and evaluation tools? If it does, which method, alternative or traditional, is predicted better by self-efficacy toward measurement and evaluation practices?

The researchers decided to perform Structural Equation Modeling (SEM) by Amos 4.0 to answer these problems. In SEM, it is possible to check the relationship between variables and confirm the theoretical structure of a scale while examining whether the model fits the data (Byrne, 2001). Moreover, Structural Equation Modeling enhances examining direct and indirect relationships between different variables (Kline, 2004).

Before running SEM, the researchers computed mean values for five factors of TEMES, year, frequency of using different measurement and evaluation tools

(separate mean scores for alternative measurement and evaluation tools and traditional ones). Then, normality of scores obtained by TEMES, TTSES, year of teaching, Traditional-ME, and Alternative-ME was examined by checking skewness/ kurtosis values, histograms with normal curves, and q-q plots. Among the skewness and kurtosis values, there were no value exceeding the range of +-3 and this was a criterion to provide normality (Tabachnick & Fidell, 2007). In addition, histograms with normal curves indicated normality and there were no cases which made the distribution non normal in q-q plots.

Table 4.8 displays zero-order correlations, mean, and standard deviation values of observed variables of structural model. There were significant correlations between the factors of the TEMES, TTSES, Alternative-ME, and Traditional-ME; but, year in teaching was not in a significant relationship with any of these variables. Among the factors of the TEMES, the fourth factor (M= 6.96, SD= 1.10) had the highest mean score. This indicated that participating teachers reported that their self-efficacy toward analyzing the results of measurement and evaluation was higher compared to their efficacy in other measurement and evaluation practices, i.e., determining assessment goals and techniques, developing assessment tools, administering assessment tools and evaluating the results, and using/ sharing the results. In addition, teachers reported that they used traditional measurement and evaluation tools (M= 3.48, SD= .68) more frequently than alternative tools (M= 2.85, SD= .84).

Table 4.8

Zero-order Correlations, Means, and Standard Deviations for Study Variables

	1	2	3	4	5	6	7	8	9
1. F_1									
2. F_2	.79*								
3. F_3	.77*	.84*							
4. F_4	.74*	.78*	.79*						
5. F_5	.69*	.71*	.79*	.69*					
6. Year	.09	.09	.10	.08	.08				
7. TTSES	.54*	.51*	.54*	.50*	.53*	.06			
8. Alternative-ME	.16*	.22*	.21*	.10*	.14*	.08	.22*		
9. Traditional-ME	.10*	.10*	.11*	.03*	.13*	-.06	.14*	.58*	
M	6.84	6.64	6.91	6.96	6.92	16.03	6.96	2.85	3.48
SD	1.02	1.09	1.07	1.10	1.10	8.41	.82	.84	.69

* $p<.05$

In consideration with the related literature, the researchers constructed a model that displays the relationship between year in teaching, frequency of using alternative (Alternative-ME) and traditional (Traditional-ME) measurement and evaluation tools, teacher self-efficacy toward measurement and evaluation practices (TEMES) and teachers' sense of efficacy (TTSES). In this study, TTSES was represented as one-dimensional manifest variable rather than latent variable with three factors (as proposed). TTSES has been used as one-dimensional in the literature as three factors are highly correlated (Deemer & Minke, 1999; Goddard, Hoy, & Woolfolk Hoy, 2000; Robert & Henson 2001). The model is illustrated in Figure 4.4. As stated before, frequency of using different measurement and evaluation tools was calculated in two distinct scores as frequency of using alternative measurement and evaluation tools (Alternative-ME) and traditional

tools (Traditional-ME); therefore, these are represented as separate variables in the model. In this model, TEMES is supposed to predict Alternative-ME and Traditional-ME, and year in teaching is supposed to directly relate to Alternative-ME, Traditional-ME, TEMES, and TTSES.

At first attempt, SEM analysis yielded χ^2 = 221.42 (p<.00), RMSEA = .15, NNFI = .97, and CFI = .99. Although NNFI and CFI values were above .90 which is a criterion for model fit, RMSEA and chi-square indicated the poor fit. Particularly RMSEA value was higher than .10, indicating poor fit (MacCallum, Browne & Sugawara, 1996). In addition, significant result of chi-square statistic displays that the specified model is different than observed data. Yet, chi-square statistic is sensitive to sample size. After this result was evaluated, the researchers examined the modification indices between error pairs and realized that errors of Alternative-ME (e_8) and Traditional-ME (e_9) had the highest covariance compared to other error pairs. Bollen (1989) suggested connecting the errors of measurement in a case of relatively high covariance between them. Considering this, the errors were connected and the model was analyzed again; because both the Alternative-ME and Traditional-ME were measuring the frequency of using different measurement and evaluation tools and both errors of measurement (e_8 and e_9) were belong to the same source, i.e., FMES (Frequency of Using Different Measurement and Evaluation Tools Scale). By this change, fit measures differed in a way that the results indicated mediocre fit. Namely, NNFI, CFI, and RMSEA values were found as follows: .99, .99, and .07, respectively. NNFI and CFI values indicated that model fit the data (Byrne, 2001), and RMSEA was found to be .07 with a 90% confidence level of .05 to .09, indicating reasonable fit (MacCallum, Browne, & Sugawara, 1996).

Unstandardized estimates are displayed in Table 4.9 and standardized estimates are shown in Figure 4.4. Teachers' sense of efficacy toward measurement and evaluation practices significantly predicted the frequency of using both alternative

(Alternative-ME) and traditional (Traditional-ME) measurement and evaluation tools. However, squared multiple correlation for Alternative-ME (.05) was higher than that of Traditional-ME (.02). That is, TEMES explained more variance in Alternative-ME. However, when direct and indirect effects were examined, year in teaching was found as a non significant predictor of teachers' sense of efficacy toward measurement and evaluation practices, Traditional-ME, Alternative-ME, and teachers' sense of efficacy. The relationship between TTSES and TEMES was also significant with a standardized estimate of .57. Five specified factor loadings in the model were also statistically significant as expected.

Table 4.9

Unstandardized Estimates for Latent and Manifest Variables

	Estimate	*SE*	*p*
Regression weights			
TTSES ← Year	.01	.01	.17
TEMES ← Year	.01	.00	.24
TEMES ← TTSES	.59	.05	.00
F_1← TEMES	1.00		
F_2← TEMES	1.33	.05	.00
F_3← TEMES	1.15	.04	.00
F_4← TEMES	1.15	.05	.00
F_5← TEMES	1.03	.05	.00
Alternative-ME ← TEMES	.19	.05	.00
Traditional-ME ← TEMES	.09	.04	.02
Alternative-ME ← Year	.01	.01	.23
Traditional-ME ← Year	-.01	.00	.15
Variances			
e_1	.69	.05	.00
e_2	.51	.05	.00
e_3	.29	.02	.00
e_4	.21	.02	.00
e_5	.15	.02	.00
e_6	.37	.03	.00
e_7	.39	.03	.00
e_8	.47	.03	.00
e_9	.68	.05	.00
Squared Multiple Correlations			
TTSES	.01		
TEMES	.33		
Alternative-ME	.05		
Traditional-ME	.02		
F_1	.73		
F_2	.82		
F_3	.87		
F_4	.73		
F_5	.68		

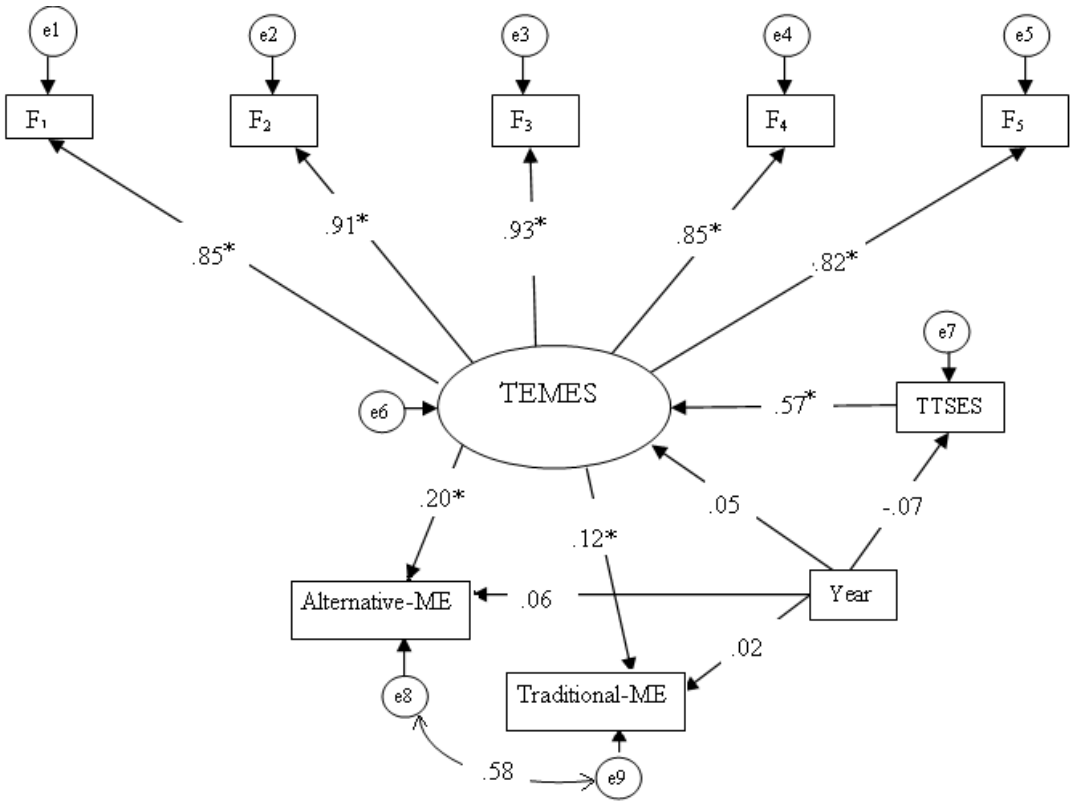

Figure 4.4 *Structural Model Representing the Relationship between Teachers' Sense of Efficacy toward Measurement and Evaluation Practices, Year, Frequency of Using Different Measurement and Evaluation Tools and Teachers' Sense of Efficacy.*
Note. TEMES stands for efficacy in measurement and evaluation practices, TTSES for teacher self-efficacy, Alternative-ME for frequency of using alternative assessment, and Traditional-ME for frequency of using traditional assessment.

4.5. Summary

In this chapter, various statistical analyses are presented in detail to verify the theoretical structure of the new instrument (TEMES), signify reliability and validity evidences for the scales administered in this study, and understand the relationships between the variables. The results of separate CFAs for TEMES and TTSES were reported with individual model representations. CFA resulted in that there were three factors of TTSES and TEMES had a 5-factor model with fit

indices indicating mediocre fit. In addition, separate one way MANOVAs yielded no significant effect of gender on the factors of TEMES, but of teaching level. The results of Canonical Correlation Analysis indicated that TEMES is a valid and reliable measure to assess teachers' sense of efficacy as well as TTSES. Lastly, a structural model was specified and evaluated in terms of fitting the data of teachers who work in public elementary and secondary schools in Ankara, Samsun, and İstanbul. SEM analysis yielded satisfactory findings. Results indicated that there was a positive correlation between teachers' sense of efficacy toward measurement and evaluation practices and the frequency of using alternative and traditional measurement and evaluation tools. However, the portion of variance explained in the frequency of using alternative measurement and evaluation tools was more than that of the traditional tools. Year of teaching was not found to be a significant predictor of TTSES, TEMES, and the frequency of using alternative and traditional measurement and evaluation tools. In the next chapter, it is discussed why the explained variance in the frequency of using alternative tools transcend the frequency of using traditional tools, together with the remaining findings.

CHAPTER V

DISCUSSION

In this chapter, purpose of the study, research design, and results of the study are summarized and discussed. In consideration with the results and shortcomings of this study, implications for practice and recommendations for further research studies are presented.

5.1. Discussion of the Study Results

This study was an associational research study in which the relationships between teacher self-efficacy toward measurement and evaluation practices, year in teaching, teachers' sense of efficacy, and frequency of using alternative and traditional measurement and evaluation tools were examined through administering a newly developed instrument. Participants were 394 teachers who work in public elementary and secondary schools in Ankara, Samsun, and İstanbul.

The purpose of this study was three-fold: (1) to develop an instrument to measure teacher self-efficacy toward measurement and evaluation practices and frequency of using different measurement and evaluation tools; (2) to evaluate the psychometric properties of the newly developed instrument; (3) to test a model which examines (a) whether year in teaching predicts teacher self-efficacy beliefs toward measurement and evaluation practices and frequency of using different measurement and evaluation tools and (b) whether teacher self-efficacy beliefs

toward measurement and evaluation practices predict frequency of using traditional and alternative measurement and evaluation tools.

The items of *Teacher Self-Efficacy toward Measurement and Evaluation Practices Scale* (TEMES) were on a 9-point rating scale ranging from "nothing" to "a great deal." Confirmatory Factor Analysis provided evidence for five-factor structure of the TEMES. The factors were: efficacy for determining assessment goals and techniques, efficacy for developing assessment tools, efficacy for administering assessment tools and evaluating the results, efficacy for analyzing the results, and efficacy for using/ sharing the results. Cronbach's alpha coefficients of these five factors were satisfactory, ranging from .76 to .87. Furthermore, investigation of the factors of the TEMES and Turkish Teachers' Sense of Efficacy Scale (TTSES) provided additional validation evidence, as the findings were consistent with the literature.

Descriptive statistics resulted in teacher self-efficacy toward measurement and evaluation practices with a mean score of 6.83 (*SD*=.98), indicating that participating teachers were "quite a bit" efficacious in measurement and evaluation practices. That is, participating teachers who work in public primary or high schools see themselves pretty good in measurement and evaluation practices. Similarly, considering the mean scores obtained from the Turkish Teachers' Sense of Efficacy Scale (TTSES), the participating teachers were efficacious in student engagement, instructional strategies, and classroom management. Contrary to this result, teachers' responses to study of MoNE, ERDHO, and The Head Office of Education and Training Board (2006) on teacher qualifications indicated that teachers did not feel comfortable with measurement and evaluation practices. Çakan (2004) concluded that most of the teachers perceived themselves as unqualified in consideration with the measurement and evaluation practices. In the current study, it appeared that participating teachers reported use of traditional

measurement and evaluation tools more than alternative measurement and evaluation tools. This finding is consistent with Çakan's study (2004), which reported that teachers have a tendency to use traditional measurement and evaluation tools.

Analysis conducted via Structural Equation Modeling indicated that year was a non significant predictor for teachers' sense of efficacy toward measurement and evaluation practices. In other words, there is no relationship between year in teaching and self-efficacy toward measurement and evaluation practices. This finding seems unexpected because Bandura (1997) proposed that the primary source of self-efficacy beliefs is mastery experiences. Mastery experiences were defined as the performances in which people act by their own. On the other hand, he noted that not the year of experience but evaluation of these experiences is important for the development of self-efficacy beliefs. The level, strength, and generality dimensions of the experiences tend to be questioned by the performers/ people to develop self-efficacy toward an action. Gür (2008), in a study with science and mathematics teachers, found that although the year of experience was not a significant predictor of teacher self-efficacy, satisfaction of the performance was statistically significant. Similar to Gür's study, our findings provided support for Bandura's assertion.

Our findings not only provided support theoretically, but also were consistent with the findings of studies conducted in Turkey. For example, Karaca (2008) concluded that teachers' perceived levels of efficacy in measurement and evaluation in education do not change significantly by years of teaching experience. In a study of teacher qualifications, MoNE, ERDHO, and Head Office of Education and Training Board (2006) reported that 70% of participating teachers (N= 2242) reported that they have difficulty in measurement and evaluation practices. In addition, there was no significant difference between

experienced and novice teachers in terms of their perceived level of qualification in measurement and evaluation practices. Similarly, in the present study, year of teaching was not a significant predictor of the frequency of using neither alternative nor traditional measurement and evaluation tools.

Finally, findings of SEM Analysis also showed that teachers' sense of efficacy toward measurement and evaluation practices predicted both using alternative and traditional measurement and evaluation tools. However, teacher self-efficacy toward measurement and evaluation practices contributed more to the explained variance in frequency of using alternative measurement and evaluation tools than frequency of using traditional tools. This finding was thought to be in an association with the literature findings in which efficacious teachers were found to be open to new ideas (Gibson & Dembo, 1984) and implementing new instructional methods rather than traditional ones (Ghaith & Yaghi, 1997). Regarding these findings, we can say that developing teacher self-efficacy toward measurement and evaluation practices has an influence on increasing the frequency of using alternative measurement and evaluation tools.

5.2. Implications for Practice

Based on the findings of the study, the following implications can be made for practice:

One of the aims of the current study was to develop a scale assessing teacher self-efficacy toward measurement and evaluation practices. Findings indicated that Teacher Self-Efficacy toward Measurement and Evaluation Practices Scale (TEMES) is a promising tool with satisfactory psychometric properties. In addition, it was found to be a multi-dimensional tool assessing five dimensions: efficacy for determining assessment goals and techniques, efficacy for developing

assessment tools, efficacy for administering assessment tools and evaluating the results, efficacy for analyzing the results, and efficacy for using/ sharing the results.

Teacher self-efficacy toward measurement and evaluation practices was not affected by year in teaching; however, it had an influence on frequency of using alternative and traditional measurement and evaluation tools. In consideration with these results, it is not realistic to expect that teachers who work for years in teaching develop higher self-efficacy toward measurement and evaluation practices compared to novice teachers. Hence year in teaching is not a significant contributor for teacher self-efficacy toward measurement and evaluation practices. As stated before, mastery experiences (own experiences of a teacher), vicarious experiences (observing performances of other teachers), social persuasion (being approved by other teachers or administrators), and physiological and emotional states (e.g., being able to cope with stress factors, enhance health functioning) are the sources to develop self-efficacy. To enhance teachers' self-efficacy in measurement and evaluation practices, teachers should be encouraged to use alternative measurement and evaluation tools by experts in measurement and evaluation practices such as members of faculties of education. Teachers' preferences on measurement and evaluation practices should be studied more and the results should be shared with in-service teachers. It may also be possible for teachers to observe each other while they are administering different measurement and evaluation tools.

Regarding the results of the study, having high self-efficacy toward measurement and evaluation practices makes teachers use more alternative measurement and evaluation tools. If teachers believe in themselves that they can use more alternative tools, they will gain experience in alternative methods. Teachers can

participate workshops, conferences or seminars which are organized by experts in the field of measurement and evaluation.

5.3. Recommendations for Further Research

In this research study, most of the items were developed based on the teacher qualifications which were proposed by MoNE. In the development process of Teacher Self-Efficacy toward Measurement and Evaluation Practices Scale (TEMES), experts' opinions were also considered. In an effort to improve the current scale, in addition to the experts' opinions, teachers' and administrators' opinions can be taken about the qualifications in measurement and evaluation practices. In addition, further validation studies are required with different populations.

Because of the restrictions in time and financial sources, the data for this study were collected in three cities of Turkey. A broader study can be useful in terms of defining Turkish teachers' self-efficacy toward measurement and evaluation practices. In order to generalize the results of the study, data can be collected from more participating teachers in further research studies. Note that the data were not collected from private school teachers due to the same restrictions. Further studies should investigate whether there is a difference between private and public school teachers in terms of their sense of efficacy beliefs toward measurement and evaluation practices. In addition, comparisons by teaching branch (i.e., teaching science versus teaching social science) were not mentioned because of unequal cell sizes. In further studies, researchers can examine this difference as well.

TEMES is an instrument to measure teacher self-efficacy toward measurement and evaluation practices. In this quantitative study, the researchers examined the relationship between year in teaching, teachers' sense of efficacy, frequency of

using alternative and traditional measurement and evaluation tools, and teacher self-efficacy toward measurement and evaluation practices. However, there are still other variables which can be important predictors for self-efficacy toward measurement and evaluation practices or which can be affected by teachers' sense of efficacy beliefs toward measurement and evaluation practices. Related literature has already indicated some significant variables which have a correlation with teacher self-efficacy; for example, enthusiasm for teaching (Guskey, 1984), student achievement (Armor et al., 1976; Ross, 1992), and planning and organization (Freidman & Kass, 2002). In future studies on teacher self-efficacy toward measurement and evaluation practices, whether there is a relationship between these variables and self-efficacy in measurement and evaluation practices can be investigated. The results will be practical in terms of understanding how to help teachers develop self-efficacy toward measurement and evaluation practices.

To examine the predictors and affecting factors of teachers' sense of efficacy towards measurement and evaluation practices, longitudinal studies can be conducted in private and public schools.

REFERENCES

Allison, P.D. (2002). *Missing data.* Thousand Oaks, CA: Sage.

Akkoyunlu, B., Orhan, F. & Umay, A., (2005). A study on developing teacher self-efficacy scale for computer teachers. *Hacettepe Universitesi Egitim Fakultesi Dergisi, 29,* 1-8.

Arbuckle, J. L. (1999). *Amos 4.0 [Computer software]*. Chicago, IL: Small Waters.

Armor, P., Conry-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., et al. (1976). *Analysis of the school preferred reading program in selected Los Angeles minority schools* (Los Angeles Unified School District). Santa Monica, CA: RAND.

Ashton, P. T., Buhr, D., & Crocker, L. (1984). Teachers' sense of efficacy: A self- or norm-referenced construct? *Florida Journal of Educational Research, 26*(1),29-41.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*,191-215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1991). Social cognitive theory of moral thought and action. In W. M. Kurtines & J. L. Gewitz (Eds.), *Handbook of moral behavior and development* (Vol. 1, pp. 45-103). Hillsdale, NJ: Erlbaum.

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of Human Behavior* (Vol. 4, pp. 71-81). New York: Academic Press. (Reprinted in H. Friedman [Ed.], *Encyclopedia of Mental Health*. San Diego: Academic Press, 1998).

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman and Company.

Bandura, A. (2006). Guide for constructing self-efficacy scales. F. Pajares & T. Urdan (Eds.). *Self-efficacy beliefs of adolescents*, 5, 307-337. Greenwich, CT: Information Age Publishing.

Bandura, A. & Cervone, D. (1986). Differential engagement and self-reactive influences in cognitive motivation. *Organizational Behavior and Human Decision Processes, 38,* 92-113.

Bandura A., Barbaranelli C., & Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development, 72,* 187-206.

Bandura, A., Caprara, G.V., Barbaranelli, C., Gerbino, M. & Pastorelli, C. (2003). Role of Affective Self-Regulatory Efficacy in Diverse Spheres of Psychosocial Functioning. *Child Development, 74(3),* 769-782.

Benabou, R. & Tirole, J. (2002). Self-Confidence and personal motivation. *The Quarterly Journal of Economics,117(3),* 871-915.

Bentler, P.M. & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 558-606.

Bentler, P.M. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 107,* 238-246.

Betz, N. E., Klein, K., & Taylor, K. (1996). Evaluation of a short form of the Career Decision Self-Efficacy Scale. *Journal of Career Assessment, 4,* 47-57.

Bıkmaz, F. (2004). Sınıf öğretmenlerinin "fen öğretiminde öz-yeterlilik inancı" ölçeğinin geçerlik ve güvenirlik çalışması. *Milli Eğitim Dergisi,* 161.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Bollen, K. A. (1989). *Structural equations with latent variables.* Wiley-Interscience Publication, USA.

Bollen, K. A. & Long, J. S. (Eds.) (1993). *Testing structural equation models.* Newbury Park, CA: Sage.

Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research and Evaluation*, 8(9). Retrieved September 12, 2009 from http://PAREonline.net/getvn.asp?v=8&n=9.

Byrne, B. M. (2001). *Structural equation modeling: Basic concepts, applications, and programming.* Lawrence Erlbaum Associates, Inc.

Cantürk-Günhan, B. & Başer, N. (2007). Geometriye yönelik öz-yeterlik ölçeğinin geliştirilmesi. *Hacettepe University Journal of Education, 33,* 68-76.

Chalmers, D.J. (1997). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press US, pp. 25-26.

Çakan, M. (2004). Öğretmenlerin ölçme-değerlendirme uygulamaları ve yeterlik düzeyleri: İlk ve ortaöğretim. *Ankara University, Journal of Faculty of Educational Sciences, 37(2),* 99-114.

Çapa, Y. (2005). *Factors influencing first-year teachers' sense of efficacy*. Unpublished doctoral dissertation, The Ohio State University.

Çapa, Y., Çakıroğlu, J. & Sarıkaya, H. (2005). The development and validation of a Turkish version of the teachers' sense of efficacy scale. *Eğitim ve Bilim, 30 (137),* 74-81.

Deemer, S. A., & Minke, K. M. (1999). An investigation of the factor structure of the Teacher Efficacy Scale. *Journal of Educational Research, 93,* 3-10.

Dellinger, A. B., Bobbettb, J.J., Olivier, D.F., & Chad D. E.(2007). Measuring teacher self-efficacy beliefs: Development and use of the TEBS-Self. *Teaching and Teacher Education*, doi:10.1016/j.tate.2007.02.010.

Dembo, M. H., & Gibson, S. (1985). Teachers' sense of efficacy: An important factor in school improvement. *The Elementary School Journal, 86,* 173-184.

DeVellis, R.F. (2003). *Scale Development: Theory and applications* (2nd Ed.). Newbury Park, CA: Sage.

Erdem, E. & Demirel, Ö. (2007). Teacher self-efficacy belief. *Social Behavior and Personality, 35 (5),* 573-586.

Erdoğan, M. (2007). Yeni öğretim programı ve yeni ölçme-değerlendirme yaklaşımları. *Eğitimde Yeni Yaklaşımlar.* ODTÜ Mezunlar Derneği Yayınları, 15-33.

Field, A. (2006). *Discovering statistics using SPSS.* Thousand Oaks. CA: Sage.

Fraenkel, J. R. ve Wallen N.E. (2008). *How to design and evaluate research in education* (7th Ed.). New York: McGraw Hill.

Friedman, I. A. (2003). Self-efficacy and burnout in teaching: The importance of interpersonal-relations efficacy. *Social Psychology of Education, 6,* 191-215.

Friedman I.A. & Kass, E. (2002). Teacher self-efficacy: A classroom organization conceptualization. *Teaching and Teacher Education 18,* 675–686.

Gelbal, S. & Kelecioglu, H. (2007). Teachers' proficiency perceptions of about the measurement and evaluation techniques and the problems they confront. *Hacettepe University Journal of Education, 33,* 135-145.

Ghaith, G. & Yaghi, H. (1997). Relationships among experience, teacher efficacy, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education, 13(4),* 451-458.

Gibson, S. & Dembo, M.H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76,* 569-582.

Gist, M. E. & Mitchell, T. R. (1992). Self-efficacy: A theoretical Analysis of its determinants and malleability. *Academy of Management: The Academy of Management Review, 17(2),* pp.183.

Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37,* 479-507.

Gravetter, F. J. & Wallnou, L.B. (2007). *Statistics for the behavioral sciences* (7th ed). Thomson Wadsworth.

Guskey, T.R. (1984). The influence of change in instructional effectiveness upon the affective characteristics of teachers. *American Educational Research Journal, 21(2),* 245-259.

Guskey, T.R. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional motivation. *Teaching and Teacher Education, 4(1),* 63-69.

Guskey, T.R. & Passaro, P. (1994). Teacher efficacy: A study of construct dimensions. *American Educational Research Journal, 31 (3),* 627-643.

Gür, G. (2008). *A study of the predictors of teachers' sense of efficacy beliefs*. Unpublished master thesis, Middle East Technical University.

Hair, J.F., Anderson, R.E., Tatham, R.L., & Black, W. C. (2006). *Multivariate Data Analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Henson, R. K. (2002). From adolescent angst to adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. *Educational Psychologist, 37(3),* 137–150.

Hoy, W. K., & Woolfolk, A. E. (1993). Teachers' sense of efficacy and the organizational health of schools. *Elementary School Journal, 93,* 335-372.

Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications.* Thousand Oaks, CA: Sage.

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

Karaca, E. (2008). A investigation of primary and high school teachers' perception levels of efficacy of measurement and evaluation in education in Turkey. *Social Behavior and Personality, 36(8), 1111-1122.*

Karadeniz, C. (2005). Developing a self-efficacy scale of geography for pre-service teachers of social sciences: Its validity and reliability. *Ondokuz Mayıs Universitesi Egitim Fakultesi Dergisi, 20,* 63-69.

Kline, R.B. (2004). *Principles and Practice of Structural Equation Modeling.* Guilford Press, 2nd Ed.

Lee, V., Dedrick, R., & Smith, J. (1991). The effect of the social organization of schools on teachers' efficacy and satisfaction. *Sociology of Education, 64,*190-208.

Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data.* New York: John Wiley.

MacCallum, R.C., Browne, M.W., & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1,* 130-149.

Meyers, L.S., Gamst, G. & Guarino, A.J. (2006). *Applied multivariate research: design and interpretation.* Thousand Oaks. CA: Sage.

MoNE & ERDHO (2004). *The qualifications for teaching*. Retrieved from http://oyegm.meb.gov.tr/yet/

MoNE, ERDHO & Head Office of Education and Training Board (2006). Ogretmen *Genel Yeterlikleri Calismasi Mevcut Durum Tespit Raporu*. Retrieved from http://otmg.meb.gov.tr/Yayin.htm.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66,* 533-578.

Riggs, I. M. & Enochs, L.G. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education, 74(6),* 625-637.

Roberts, J. K., & Henson, R. K. (2001). *A confirmatory factor analysis of a new measure of teacher efficacy: Ohio State Teacher Efficacy Scale*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Rose, J.S. & Medway, F.J. (1981). Measurement of teachers' beliefs in their control over student overcome. *Journal of Educational Research, 74,* 185-190.

Ross, A. J. (1992). Teacher efficacy and the effects of coaching on student achievement. *Canadian Journal of Education, 17(1),* 51-65.

Ross, J.A., Cousins, J.B., & Gadala, T. (1996). Within-teacher predictors of teacher efficacy. *Teaching & Teacher Education, 12(4),* 385-400.

Ross, J. A., Cousins, J. B., Gadalla T., & Hannay, L. (1999). *The effects of course assignment on teacher efficacy in restructuring secondary schools*. Paper Presented at the Annual Meeting of the American Educational Research Association Montreal, Canada. (ERIC Document Reproduction Service No. ED430280).

Rotter, J.B. (1966). Generalized expectancies for intent versus external control of reinforcement. *Psychological Monographs, 80*, 1-28.

Rychtarik, R. G., Prue, D. M., Rapp, S. R., & King, A. C. (1992). Self-efficacy, aftercare and relapse in a treatment program from alcoholics. *Journal of Studies on Alcohol, 53*, 435-440.

Schwarzer, R. & Schmitz, G.S. (2005). *Perceived self-efficacy and teacher burnout: a longitudinal study in ten schools*. Research paper. Freie Universitat Berlin, Germany.

Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors*. Paper presented at the Psychometric Society Annual Meeting, Iowa City, IA.

Stevens, J.P. (2002). *Applied multivariate statistics for the social sciences* (4th ed). Hillsdale, NJ: Erlbaum.

Tabachnick, B.G. & Fidell, L.S. (2007). *Using Multivariate Statistics*. Pearson Education Company, 5th Ed.

Thatcher, J.B. & Perrewé, P.L (2002). An empirical examination of individual traits as antecedents to computer anxiety and computer self-efficacy. *MIS Quarterly, 26(4),* 381-396.

Tschannen -Moran, M., Woolfolk Hoy, A., & Hoy, W. K., (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68,* 202-248.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing and elusive construct. *Teaching and Teacher Education, 17 (7),* 783-805.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education, 23,* 944-956.

Ware, H. & Kitsantas, A.(2007). Teacher and collective efficacy belief as predictors of professional commitment. *The Journal of Educational Research, 100(5),* 303-310.

Woolfolk Hoy, A. & Spero, R.B., (2005). Changes in teacher efficacy during early years of teaching: A comparison of four measures. *Teaching and Teacher Education, 21,* 343-356.

Vitali, G. (1993). *Factors influencing teachers' assessment and instructional practices in an assessment-driven educational reform*. Unpublished doctoral dissertation, University of Kentucky, Lexington, KY.

Yılmaz, M., Koseoglu, P., Gercek, C. & Soran, H., (2004). Adaptation of a teacher self-efficacy scale to Turkish. *Hacettepe Universitesi Egitim Fakultesi Dergisi, 27*, 260-267.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation*. (pp. 13-39). San Diego, CA: Academic Press.

APPENDICES

APPENDIX A

Öğretmen Özyeterlikleri Araştırması

Bu çalışma, öğretmenlerin ölçme ve değerlendirme uygulamalarında kendi özyeterlik inançlarını ölçmeye yöneliktir. Aşağıdaki anket öğretmenlerin bu alandaki özyeterlik inançlarını belirlemek amacıyla oluşturulmuştur. Anket 3 bölümden oluşmaktadır.

Bu çalışmada, katılımcıların isim, adres, telefon vb. hiçbir kişisel bilgisi yer almayacaktır. Verdiğiniz cevaplar, araştırma süresinde ve sonrasında saklı tutulacaktır. Ancak aşağıda yer alan ve araştırmada belirtilmesi gereken yaş, cinsiyet, meslekteki yılı, branş ve öğretim kademesi bilgileri kesinlikle doldurulmalıdır.
Bu bilgiler, araştırma katılımcılarının karakterinin belirlenmesi açısından büyük önem taşımaktadır.

FORM ID

Katkılarınız için teşekkür ederim.

Fatma Rana Ceylandağ
ODTÜ Sosyal Bilimler Enstitüsü
Eğitim Bilimleri Yükseklisans Öğrencisi

Demografik Bilgiler

Lütfen, aşağıdaki soruları cevaplandırınız ve yuvarlakla belirlenmiş kısımlarda size uygun olanını işaretleyiniz.

1. **Öğrenim durumunuz (en son aldığınız diploma derecesi) nedir?**
   - (1) Öğretmen okulu- Yüksek öğretmen okulu
   - (2) Ön Lisans
   - (3) Lisans
   - (4) Yüksek Lisans
   - (5) Doktora

2. **Üniversite eğitimi gördüğünüz fakülteninadı :**
   _____

3. **Üniversite eğitimi gördüğünüz anabilim dalının adı:**
   _____

4. **Yaşınız:** _____

5. **Cinsiyetiniz nedir?**
   - (1) Bayan
   - (2) Bay

6. **Kaç yıldır öğretmenlik yapmaktasınız?** _____(yıl)

7. **Hangi branşta öğretmenlik yapmaktasınız?**
   - (1) Türkçe ( )
   - (2) Fen Bilgisi ( )
   - (3) Fizik ( )
   - (4) Kimya ( )
   - (5) Biyoloji ( )
   - (6) Matematik ( )
   - (7) Sosyal Bilgiler ( )
   - (8) İngilizce ( )
   - (9) Diğer ( )

8. **Hangi öğretim kademesinde öğretmenlik yapmaktasınız?**
   - (1) İlköğretim 1. Kademe (1-5. sınıflar)
   - (2) İlköğretim 2. Kademe (6-8. sınıflar)
   - (3) Ortaöğretim (Lise 1, 2, 3, 4)

9. **Öğretmenlik yaptığınız okul türü nedir?**
   - (1) Düz Lise
   - (2) İlköğretim Okulu Devlet
   - (3) Özel Lise
   - (4) İlköğretim Okulu Özel
   - (5) Anadolu Lisesi
   - (6) Diğer

10. **Okulunuzda ölçme ve değerlendirme birimi var mı?**
    - (1) Evet
    - (2) Hayır

11. **Üniversite eğitiminiz sırasında ölçme ve değerlendirmeye yönelik bir ders aldınız mı?**
    - (1) Evet
    - (2) Hayır

12. **Ölçme ve değerlendirme alanına yönelik hizmet içi eğitim aldınız mı?**
    - (1) Evet
    - (2) Hayır

APPENDIX B

I. Bölüm:

Ankette yer alan soruların cevap seçenekleri olarak 1 ile 9 arası derecelendirme yapılmıştır. Rakamların karşılık geldiği anlamlar şöyle verilmiştir:

1= yetersiz, 3= çok az yeterli, 5= biraz yeterli, 7= oldukça yeterli, 9= çok yeterli.

Lütfen sorulara cevap verirken 1 ile 9 arasındaki derecelendirmede size uygun olan rakamı yuvarlak içine alınız.

		Yetersiz		Çok az yeterli		Biraz yeterli		Oldukça yeterli		Çok Yeterli
1.	Değerlendirme amacınızı (biçimlendirmeye yönelik, sonlandırmaya yönelik vb.) ne kadar iyi belirleyebilirsiniz?	1	2	3	4	5	6	7	8	9
2.	Eğitim programında (müfredat) önerilen hedef ve kazanımlara uygun ölçme ve değerlendirme araçlarını belirlemede ne kadar başarılısınız?	1	2	3	4	5	6	7	8	9
3.	Hazırlanan soruların öğrenilecek içeriğe uygun olmasını sağlamada ne kadar iyisiniz?	1	2	3	4	5	6	7	8	9
4.	Öğrencinin ilgi ve ihtiyaçları doğrultusunda çeşitli ölçme ve değerlendirme araçları geliştirmede ne kadar başarılısınız?	1	2	3	4	5	6	7	8	9
5.	Öğrencilerin performans ve gelişim düzeylerini düzenli olarak ölçmede ne kadar iyisiniz?	1	2	3	4	5	6	7	8	9
6.	Çok yönlü değerlendirme için alternatif ölçme araçlarını (örn. portfolyo, kavram haritaları vb.) ne kadar iyi belirleyebilirsiniz?	1	2	3	4	5	6	7	8	9
7.	Ölçme ve değerlendirme sonuçlarını daha sonraki ders planlarında ne kadar etkili kullanabilirsiniz?	1	2	3	4	5	6	7	8	9
8.	Ölçme aracı geliştirmede ne kadar başarılısınız?	1	2	3	4	5	6	7	8	9
9.	Ölçme aracının geçerlilik ve güvenirliğini test etmede ne kadar başarılısınız?	1	2	3	4	5	6	7	8	9
10.	Ölçme aracını uygulamada ne kadar başarılısınız?	1	2	3	4	5	6	7	8	9
11.	Öğrencinin çalışmalarını (proje, ödev, vb.) puanlamada ne kadar başarılısınız?	1	2	3	4	5	6	7	8	9
12.	Bireysel ölçme ve değerlendirme etkinlikleri (örn. performans değerlendirme, proje) hazırlamada ne kadar iyisiniz?	1	2	3	4	5	6	7	8	9
13.	Öğrencilerinizin iyi bir şekilde değerlendirilmesine olanak sağlayacak soruları ne ölçüde hazırlayabilirsiniz?	1	2	3	4	5	6	7	8	9
14.	Bireysel farklılıkları dikkate alarak farklı ölçme ve değerlendirme araçlarını ne kadar iyi geliştirebilirsiniz?	1	2	3	4	5	6	7	8	9
15.	Hazırlanan sınav sorularındaki mevcut problemleri teşhis etmede ne kadar başarılısınız?	1	2	3	4	5	6	7	8	9
16.	Sınav sorularını iyileştirmede ne kadar başarılısınız?	1	2	3	4	5	6	7	8	9
17.	Eğitim programında (müfredat) önerilen hedef ve kazanımlara uygun ölçme araçları geliştirmede ne kadar iyisiniz?	1	2	3	4	5	6	7	8	9
18.	Ölçme sonuçlarını tablo ve grafik türü görsel biçimlere dönüştürmede ne kadar başarılısınız?	1	2	3	4	5	6	7	8	9
19.	Ölçme sonuçlarını yorumlamada ne kadar başarılısınız?	1	2	3	4	5	6	7	8	9
20.	Ölçme ve değerlendirme sonuçlarından öğrencilerinize geribildirim sağlamada ne kadar iyisiniz?	1	2	3	4	5	6	7	8	9
21.	Ölçme ve değerlendirme sonuçlarını veliler, okul yönetimi ve diğer eğitimcilerle paylaşmada ne kadar iyisiniz?	1	2	3	4	5	6	7	8	9
22.	Alternatif ölçme araçlarını (örn. kavram haritaları, yapılandırılmış grid vb) ne kadar iyi geliştirebilirsiniz?	1	2	3	4	5	6	7	8	9
23.	Öğrettiklerinizin öğrenciler tarafından kavranıp kavranmadığını değerlendirmede ne kadar iyisiniz?	1	2	3	4	5	6	7	8	9
24.	Kendi öğretim stratejilerinizi değerlendirirken öğrenci değerlendirmelerini kullanmada ne kadar iyisiniz?	1	2	3	4	5	6	7	8	9

APPENDIX C

II. Bölüm:

Bu bölümde, çeşitli ölçme ve değerlendirme araçları aşağıdaki gibi listelenmiştir.

Bu araçları ne sıklıkta kullandığınızı belirtmeniz istenmektedir. Lütfen, sizin için uygun olan seçeneği işaretleyiniz.

	Hiç	Nadiren	Ara Sıra	Sıklıkla	Her zaman	Bilgim yok
1) Yazılı sınav (uzun cevaplı yoklamalar)	1	2	3	4	5	6
2) Yazılı sınav (kısa cevaplı yoklamalar)	1	2	3	4	5	6
3) Çoktan seçmeli test	1	2	3	4	5	6
4) Kelime ilişkilendirme	1	2	3	4	5	6
5) Yazılı raporlar	1	2	3	4	5	6
6) Görüşme ve gözlem	1	2	3	4	5	6
7) Doğru-yanlış soruları	1	2	3	4	5	6
8) Eşleştirme soruları	1	2	3	4	5	6
9) Boşluk doldurma soruları	1	2	3	4	5	6
10) Drama	1	2	3	4	5	6
11) Soru-cevap	1	2	3	4	5	6
12) Portfolyo (öğrenci ürün seçki dosyası)	1	2	3	4	5	6
13) Kavram haritası	1	2	3	4	5	6
14) Yapılandırılmış grid	1	2	3	4	5	6
15) Performans değerlendirme	1	2	3	4	5	6
16) Kendi kendini değerlendirme	1	2	3	4	5	6
17) Akran değerlendirmesi	1	2	3	4	5	6

APPENDIX D

III. Bölüm:

Lütfen sorulara cevap verirken 1 ile 9 arasındaki derecelendirmede size uygun olan rakamı yuvarlak içine alınız.

		Yetersiz	Çok az yeterli	Biraz yeterli	Oldukça yeterli	Çok Yeterli
1.	Çalışması zor öğrencilere ulaşmayı ne kadar başarabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
2.	Öğrencilerin eleştirel düşünmelerini ne kadar sağlayabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
3.	Sınıfta dersi olumsuz yönde etkileyen davranışları kontrol etmeyi ne kadar sağlayabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
4.	Derslere az ilgi gösteren öğrencileri motive etmeyi ne kadar sağlayabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
5.	Öğrenci davranışlarıyla ilgili beklentilerinizi ne kadar açık ortaya koyabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
6.	Öğrencileri okulda başarılı olabileceklerine inandırmayı ne kadar sağlayabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
7.	Öğrencilerin zor sorularına ne kadar iyi cevap verebilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
8.	Sınıfta yapılan etkinliklerin düzenli yürümesini ne kadar iyi sağlayabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
9.	Öğrencilerin öğrenmeye değer vermelerini ne kadar sağlayabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
10.	Öğrettiklerinizin öğrenciler tarafından kavranıp kavranmadığını ne kadar iyi değerlendirebilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
11.	Öğrencilerinizi iyi bir şekilde değerlendirmesine olanak sağlayacak soruları ne ölçüde hazırlayabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
12.	Öğrencilerin yaratıcılığının gelişmesine ne kadar yardımcı olabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
13.	Öğrencilerin sınıf kurallarına uymalarını ne kadar sağlayabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
14.	Başarısız bir öğrencinin dersi daha iyi anlamasını ne kadar sağlayabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
15.	Dersi olumsuz yönde etkileyen ya da derste gürültü yapan öğrencileri ne kadar yatıştırabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
16.	Farklı öğrenci gruplarına uygun sınıf yönetim sistemi ne kadar iyi oluşturabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
17.	Derslerin her bir öğrencinin seviyesine uygun olmasını ne kadar sağlayabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
18.	Farklı değerlendirme yöntemlerini ne kadar kullanabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
19.	Birkaç problemli öğrencinin derse zarar vermesini ne kadar iyi engelleyebilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
20.	Öğrencilerin kafası karıştığında ne kadar alternatif açıklama ya da örnek sağlayabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
21.	Sizi hiçe sayan davranışlar gösteren öğrencilerle ne kadar iyi baş edebilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
22.	Çocuklarının okulda başarılı olmalarına yardımcı olmaları için ailelere ne kadar destek olabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
23.	Sınıfta farklı öğretim yöntemlerini ne kadar iyi uygulayabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				
24.	Çok yetenekli öğrencilere uygun öğrenme ortamını ne kadar sağlıyabilirsiniz?	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨				