SIXTH, SEVENTH AND EIGHTH GRADE TEACHERS' CONCEPTION OF ASSESSMENT

A THESIS SUBMITTED TO
THE GRADUTE SCHOOL OF SOCIAL SCIENCES
OF MIDDLE EAST TECHNICAL UNIVERSITY

BY

EMİNE VARDAR

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

JUNE 2010

Approval of the Graduate School of Social Sciences.

_____

Assist. Prof. Dr. Aykan ERDEMİR
Director

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

_____

Prof. Dr. Ali YILDIRIM
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.

_____

Prof. Dr. Meral AKSU
Supervisor

Examining Committee Members

Prof. Dr. Meral AKSU (METU, EDS) _____
Assoc. Prof. Dr. Semra SUNGUR (METU, ELE) _____
Assist. Prof. Dr. Yeşim ÇAPA AYDIN (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.

Name, Last name: Emine Vardar

Signature :

ABSTRACT

SIXTH, SEVENTH AND EIGHTH GRADE TEACHERS' CONCEPTION OF ASSESSMENT

Vardar, Emine

M.S., Department of Educational Sciences

Supervisor: Prof. Dr. Meral Aksu

June 2010, 93 pages

The main purpose of this study was to investigate sixth, seventh, and eight grade Turkish, English, Mathematics, Science and Technology, and Social Studies teachers' conceptions of assessment. It was also aimed at finding out the differences, if any, in teachers' conceptions of assessment according to their teaching subject, years of teaching experience, in-service training, and the undergraduate institution they graduated from.

There were 414 teachers included in the study. Data were gathered through the short version of Teacher Conceptions of Assessment Scale (TCoA-IIIA), which had 27 items. Exploratory Factor Analysis results revealed that there were four factors in the scale which were called Student Accountability, School Accountability, Improvement, and Irrelevance like in the original scale.

Then, the correlation results revealed that Student Accountability, School Accountability and Improvement conceptions were significantly correlated with each other at moderate level. However, Irrelevance conception did not have a significant relationship with other conceptions.

The findings of Multivariate Analysis of Variance (MANOVA) indicated that teaching subject and in-service training did not make any significant difference in teachers' conceptions of assessment. However, years of teaching experience and undergraduate institution teachers graduated did significant differences in teachers' conceptions of assessment.

Key Words: Conception, Assessment, Measurement, Evaluation.

ÖZ

ALTINCI, YEDİNCİ VE SEKİZİNCİ SINIF ÖĞRETMENLERİNİN ÖLÇME VE DEĞERLENDİRME SÜRECİNE İLİŞKİN KAVRAYIŞLARI

Vardar, Emine

Yüksek Lisans, Eğitim Bilimleri Bölümü

Tez Yönetici: Prof. Dr. Meral Aksu

Haziran 2010, 93 sayfa

Bu araştırmanın temel amacı, ilköğretimde çalışan altıncı, yedinci ve sekizinci sınıf Türkçe, İngilizce, Matematik, Fen ve Teknoloji ve Sosyal Bilgiler öğretmenlerinin ölçme ve değerlendirme sürecine ilişkin kavrayışlarının belirlenmesidir. Ayrıca, bu araştırmayla öğretmenlerin branşı, meslekteki hizmet yılları, hizmet-içi eğitim alma durumları ve mezun oldukları yükseköğretim kurumuna göre öğretmenlerin ölçme ve değerlendirme sürecine ilişkin kavrayışları arasında fark olup olmadığı araştırılmıştır.

Araştırmaya toplam 414 branş öğretmeni katılmıştır. Veri toplama sürecinde, araştırmacılar tarafından Türkçe'ye uyarlanan "Öğretmenlerin Ölçme ve Değerlendirme Sürecine İlişkin Kavrayışları" ölçeğinin 27 maddelik kısaltılmış hali kullanılmıştır. Açıklayıcı Faktör Analizi sonuçları ölçeğin aslına benzer yapıda olduğunu ve dört faktörden oluştuğunu göstermektedir. Faktör isimleri ölçeğin aslına uygun bir şekilde "Öğrenci Sorumluğu", "Okul Sorumluluğu", "Gelişim" ve "Önemsizlik" olarak isimlendirilmiştir.

Korelasyon analizi bulguları, Öğrenci Sorumluluğu, Okul Sorumluluğu ve Gelişim boyutlarının birbirleriyle orta düzeyde anlamlı birer ilişki oluşturduğunu göstermektedir.

Verilerin analizinde Çoklu Varyans Analizi (MANOVA) kullanılmıştır. Öğretmenlerin okuttukları branş ve hizmet-içi eğitim alma durumları, öğretmenlerin ölçme ve değerlendirme süreci hakkındaki kavrayışları üzerinde anlamlı bir farklılığa yol açmamıştır. Ancak, öğretmenlerin meslekteki hizmet yılları ve öğrencilerin mezun oldukları yükseköğretim kurumları, öğretmenlerin ölçme ve değerlendirme süreci hakkındaki kavrayışları üzerinde anlamlı farklılığa yol açmıştır.

Anahtar Kelimeler: Kavrayış, Ölçme ve Değerlendirme, Ölçme, Değerlendirme.

To my handsome nephew Burak

and

my cutest nieces Gülce and Ece…

ACKNOWLEDGMENTS

Firstly, I would love to thank to my advisor Prof. Dr. Meral Aksu for her wonderful guidance and effective feedback on each manuscript of the parts of this thesis. Although she had various responsibilities as Dean of Faculty of Education, she was always ready to answer my calls and share long hours to work on this study.

Secondly, I am grateful to Assoc. Prof. Dr. Adnan Kan, who was a kind of co-advisor for me from Gazi University. I appreciate his patience in answering my endless questions and his unconditional help in this journey.

I would like to thank to my instructors Dr. Finlay McQuade (Emeritus) and Prof. Dr. Giray Berberoğlu who made me see the significance of assessment in teaching and learning environment. Then, I want to send my thanks to my teachers Ayse -Erdoğan Köse, Süleyman Yaman, and Fatma Çetin who helped me to shape my future by taking into account all my strengths and weaknesses as a small child in their classrooms.

I would like to thank to Prof. Dr. M. Çağatay Özdemir and Prof. Dr. Ülker Akkutay from Gazi University for their support and understanding in this process.

I am thankful to my friends Münevver Aygenç, Esra Debreli, Nilüfer Korkmaz, Vildan Akıncılar-Çal, Mehtap İnce, Canan Mesutoğlu, A. Şenay Dinçer, Erkan Koç, Sezen Keser, Işıl İşler, Subhan Ekşioğlu and Emel Tüzel for their support and friendship throughout this study.

Later, I have to thank to all teachers anonymously responded to the questionnaires. I really appreciate their help in this study.

I would like thank to Gavin T.L. Brown for his permission to adapt his scale, for his suggestions and wonderful patience to answer my e-mails.

Besides my thesis supervisor, I am also grateful to other jury members, Assist. Prof. Dr. Yeşim Çapa Aydın and Assoc. Prof. Dr. Semra Sungur, for their constructive and effective feedback that increased the quality of this study.

And, my family. I know you were always with me in this long and stressful process. Special thanks to your unconditional love and trust on me. I miss you so much…

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

AMOS: Analysis Moments of Structures

MoNE: Turkish Ministry of National Education

ERDHO/ EARGED: Educational Research and Development Head Office

HSEC: METU Human Subjects Ethics Committee

TCoA-IIIA: Teacher Conceptions of Assessment Abridged Scale

STACC: Student Accountability

SCACC: School Accountability

IMP: Improvement

IRR: Irrelevance

LDE/SBS: Level Determination Exam

KMO: Kaiser-Meyer- Olkin Measure of Sampling Adequacy

N= Sample Size

f= Frequency

M= Mean

SD= Standard Deviation

p= Significance Level

EFA= Exploratory Factor Analysis

CFA= Confirmatory Factor Analysis

ANOVA= Analysis of Variance

MANOVA: Multivariate Analysis of Variance

PASW= Predictive Analysis Software

CHAPTER I

INTRODUCTION

In this part, the background of the study, purpose, significance, and some important definitions used in the study were mentioned in order to give a general idea about the structure of the thesis.

1.1 Background of the Study

Assessment was considered to play a critical role in education, for both policy makers and practitioners by serving both accountability (how well students have learned) and instruction (how to promote higher levels of learning) purposes in educational context (Danielson, 2008). The purposes of classroom assessment are grouped as being formative, diagnostic or summative ones which are assumed to shape the classroom assessment by taking into account all assessment users' views in decision making process. Formative assessment is defined as 'all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged' (Black & William, 1998, p. 7). In summative assessment, the products are assessed by *internally* or *externally* (Harlen, 2008) in educational system. The former one includes grading, records keeping which are used to inform parents or students themselves; however, the latter serves for selection or certification through any kind of exams or vocational purposes.

1.1.1 The Educational Context in Turkey

In any educational reform done around the world, the policies including curriculum, assessment, and professional development were manipulated by the reformers; and it was commonly believed that manipulating all these factors will change teachers' practices which will result in improving instruction and increasing students' performance in this process (Cohen & Hill, 2000).

The similar policies were also practiced by Turkish Ministry of National Education (MoNE) in order to follow the technological and scientific global changes. In this process, committees in five subjects including Turkish, Mathematics, Science and Technology, Social Studies and Life Sciences started to work on curriculum revisions to have a better nationwide educational system in 2003. Each committee had curriculum development specialists, measurement and evaluation specialists, teachers and academicians. In 2004-2005 academic year, these new curricula were piloted and in following academic years, from 1st to 5th, and then from 6th to 8th grade curricula were started to be implemented nationwide.

These curriculum revisions, which were based on social, individual, economical, historical and cultural fundamentals (Koç, Isıksal, & Bulut, 2007), made some changes in teachers' and students' roles, which were based on the requirements of the so called constructivist approach that guided teachers to behave as a facilitator in teaching and learning environment by making students more active for their learning (MEB, 2005). Since these revisions made some changes in course contents, teaching strategies, materials, and measurement and evaluation techniques (Gelbal & Kelecioğlu, 2007), MoNE also made some revisions in the structure of exam system

for entrance to the high schools and introduced an exam, which was called Level Determination Exam (SBS). Actually, this exam was prepared to report mainly how much of the objectives of each curriculum (Turkish, English, Science and Technology, Social Studies, and Mathematics) was succeeded by students and then to give feedback to the system.

Lastly, it is important to understand how teachers conceive of assessment, how they use the assessment techniques, and what they need in this process by focusing on their needs and support their classroom assessment practices by preparing effective and high quality in-service trainings or professional development programs (Brown, 2002) in order to increase the quality of teaching and learning.

1.2 Purpose of the Study

This study aimed at investigating sixth, seventh, and eighth grade teachers' conceptions of assessment. Moreover, this study was conducted to examine the differences, if any, in teachers' conceptions of assessment according to their teaching subject, years of teaching experience, in-service training, and the undergraduate institution that teachers graduated from.

The idea of studying teachers' conceptions originated from the necessity of learning what practitioners (both teachers, students, parents and policy-makers) believe assessment means to them rather than what is suggested or mentioned in schools, books or in-service training or seminars about assessment. Since there was a limited body of research in Turkey to investigate teachers' conceptions of assessment,

Teacher Conceptions of Assessment Abridged Scale (TCoA –IIIA ; Brown, 2008) was adapted from English to Turkish.

1.3 Significance of the Study

The issue of how teachers' conceive of assessment was ignored and was not studied in detail to have a general understanding about teachers' conceptions of assessment. This study was supposed to contribute to the literature by investigating sixth, seventh, and eighth grade teachers' conceptions of assessment in Turkey. Since there was not any Turkish scale examining the conceptions of assessment, the researchers adapted and validated a data gathering tool into Turkish from English. According to the findings of this study, there might be effective implications for MoNE to design high quality in-service programs by focusing on the teachers' needs and weaknesses in assessment process. Further, the findings related with the variables as teaching subject, years of teaching experience, in-service training, and undergraduate institution that teachers graduated from will make concrete the issue from this perspective.

1.4 Definitions of the Terms

Conception: It involves the general mental structures, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like (Thompson, 1992, p. 141).

Assessment: A broad term meaning a process for obtaining information that is used for making decisions about students; curricula, programs, and schools; and educational policy (Brookhart & Nitko, 2008, p. 4).

Measurement: A process of gathering data that provides a more precise and objective appraisal of learning outcomes that could be accomplished by less formal and systematic procedures (Payne, 2003, p.9).

Evaluation: The process of making a value judgment about the worth of a student's product or performance (Brookhart & Nitko, 2008, p. 5).

CHAPTER II

LITERATURE REVIEW

This chapter included the related literature by referring to the leading studies in abroad and Turkey. The significance of studying teachers' conceptions and their role in teacher change were discussed in the first two sections; and then teachers' conceptions of assessment were introduced in detail by referring to four main conceptions of assessment. In final part, the studies from abroad and Turkey were discussed in order to clarify the issue in a more understandable way.

2.1 The Significance of Studying Teachers' Conceptions

What people believe or have in mind and what kind of effects of these beliefs have on their behaviors or actions were discussed in various studies. Sigel (1985) defined beliefs as 'mental constructions of experience- often condensed and integrated into schemata or concepts' (cited in Pajares 1992, p. 313) which are used to shape the behaviors. Although there is not a correct definition for beliefs, most of the researchers are explaining their own definition by focusing on their needs, mental structures and understanding by using their own experiences. In this study, the use of the term of "conception" has taken its roots from Thompson's (1992) and Brown's (2002, 2004, 2008) definitions. In the beginning of 90s, Thompson (1992) explained the distinction between teachers' beliefs and conceptions as "conceptions are the mental structures, encompassing both beliefs and any aspect of the teachers' knowledge that bears on their experience, such as meanings, concepts, propositions, rules, mental images, and the like – instead of simply teachers' beliefs" (p.141). In

addition, Brown (2008) mentioned about both affective and evaluative components of the conceptions by giving a simple example in his book: He stated that although "school" is used as a "concept" in daily life, "school is good" represents a conception in which the value, worth or purpose of school is mentioned (p. 9).

Lerman (2001) emphasized that the teachers' beliefs and practices are not separated and they are needed to be seen as a whole including both cognitive and emotional parts of the subjects if there is an expected change in teachers teaching in a more positive direction.

2.2 Examining Teachers' Conceptions in Teacher Change

It is important to involve teachers in change process in order to increase their willingness to change and to modify the change process to suit their own context (Weeden, Winter, & Broadfoot, 2002). Further, this chance given to teachers to face and reflect on their conceptions effectively will raise the awareness of teachers towards their misconceptions, misunderstandings or misuses in teaching and learning environment.

Professional development programs are considered as systematic efforts to bring about a change in the classroom practices of teachers, in their attitudes and beliefs, and in the learning outcomes of students (Guskey, 2002). Furthermore, the professional development model proposed by Flexer, Cumbo, Borko, Mayfield and Marion (1994) might be considered as an example to reveal the importance of teacher conceptions of assessment in classroom. The authors of this model mentioned that there were three key sets of beliefs impact on classroom practice

including: (a) beliefs about how children learn; (b) beliefs about what [mathematics] is important to teach children; and (c) beliefs about instruction and assessment. In their study, teachers appeared to change their beliefs as a consequence of successfully using new assessment approaches (cited in Anderson, White, & Sullivan, 2005).

Finally, in order to better discuss the teachers' conception of assessment, the next part is arranged to give detailed information about these conceptions, which are considered as one of the significant factors in having better assessment implications.

2.3. Conceptions of Assessment

While teachers are assessing the students in their classroom, their individual experiences and conceptions are affecting their students' learning and classroom performance. From this perspective, studying conceptions of assessment attracted researchers attention in last decades in order to figure out how assessment users conceive of assessment in teaching and learning environment.

The book, *Knowing What Students Know,* explains the reason of why studying the conceptions of assessment started to become popular in the literature. As stated there, since policy makers needed a tool for evidence of the accountability purposes in decision-making process, assessment is used to serve these purposes by focusing on the assessment users' conceptions (NCR, 2001).

Accountability conception, which makes teachers accountable to assign a grade to students and determine the quality of schools was dominating the educational context for a long time. In last decades, changing educational approaches from teacher-

centered to a more learner-centered learning environment made assessment focus overall process rather than the product only, therefore the improvement of students' learning and helping teachers to improve their teaching according to the assessment findings gained popularity in the world in order to increase students' deep learning.

In addition to accountability of schools and students and improvement conceptions, some qualitative studies revealed another assessment conception, which made teachers conceive of assessment as irrelevant and meaningless in their classrooms.

From New Zealand context, teacher conceptions about assessment were discussed several times by G.T.L Brown who developed a scale (TCoA III) and proposed a four-facet model for teacher conceptions of assessment (see Brown, 2002; 2004; and 2008). In his model whose components were based on the findings of a detailed and careful literature review and of the results of previous studies (TCoA I and TCoA II), four main conceptions for how teachers conceive of assessment were proposed and discussed in detail. Although, there might be various conceptions of assessment, the following parts of the review will discuss teachers' conceptions of assessment by using the similar classification aligned with the assumptions proposed in Brown (2002) as (1) assessment is useful in improving teacher instruction and student learning by providing quality information for decision-making, (2) assessment is about accountability of students through certification processes, (3) teachers or schools are made accountable thorough internal and external evaluations, and (4) assessment is irrelevant or pernicious to the work of teachers and the life of students (p. 25).

2.3.1. Improvement Conception

In assessment process, when the needs of students in a real teaching and learning environment are examined critically, it is highly recommended making students improve their higher-order thinking skills, which includes *concept understanding, rule-governed thinking, problem solving, and critical thinking* in teaching and learning environment (see Brookhart & Nitko, 2009; Musial et al., 2009; Marzano, et al., 1988).

The effect of assessment on improving teaching and learning started to become as one of the most discussed topics in the literature after the results of some international assessments were announced and some countries started to revise their educational systems through educational reforms.

It was revealed that the *assessment, instruction and teaching are inseparable*, (Frederickson& Collins, 1989 cited in Ginsburg, Jacobs, & Lopez, 1993) and therefore assessment should be carefully planned to increase learning of students (Dochy, Gijbels, & Segers, 2006). However, before increasing popularity of formative assessment in classroom environment, the assessment and instruction were considered as separate activities (Birenbaum, 2003; Struyf, Vandenberghe, & Lens, 2001) where assessment was coming at the end of learning process (Dochy, Gijbels, & Segers, 2006).

In order to have an assessment system which is successfully aligned with curriculum, teaching and learning, the data gathered in this process should be valid and reliable in order to increase the positive effects on improving teaching, students' learning and

understanding in this process. Also, these positive results for student learning provide further evidence that teachers' interpretation and uses of assessment results were valid if classroom assessment samples the most important learning targets (Brookhart & Nitko, 2009).

2.3.2 School Accountability Conception

The history of the term "accountability" or "educational accountability" goes back to the years after World War II in the United States. Especially, when the number of functionally illiterate people started to increase, people started to wonder the how well schools were performing (Wickline, 1971). Because the term accountability concerned about the ability of the educational system to prepare their citizens to meet successfully the challenges of a global economy (Bennett & Gitomer, 2009), all of the changes in society, economics, and technology impacted on the ideas about the appropriate outcomes of schooling.

It will be helpful to introduce the effects of high-stakes and low-stakes assessments on accountability in order to draw a more concrete frame for the terms of assessment and accountability. As Brown (2008) mentions that low-stakes school accountability consequences, which include classroom assessments rather than standardized nation-wide examinations or assessments, are "seen as one of the important components of any national testing system intended to fulfill both the accountability and improvement purposes" (p.20). Simply, Madaus (1988) mentioned about the properties of the regimes including high-stakes accountability consequences as: (1) generates clear and uniform standards; (2) generates easier and more objective

accountability at all levels and (3) provides concrete information on how well schools are doing for the public (cited in Weeden et al., 2002).

In order to enlarge the accountability conception and how assessment is being responsible for the accountability; Linn (2000) made detailed inferences about how high-stakes accountability positively influences and improves the instruction and student learning when teachers and students are held accountable for the results of these assessments (p.4). Moreover, Linn (2000) also explained the reasons of why policymakers have seen the assessment as one of the key components of increasing accountability as follow: (1) tests and assessments are relatively inexpensive, (2) testing and assessment can be externally mandate, (3) testing and assessment changes can be rapidly implemented, (4) results are visible (p. 4).

However, when the negative arguments about accountability were analyzed, there were some researchers who supported that students' learning is being left out of the definition and then been underused in the classroom (Sangster& Overall, 2006). Further, Bennett and Gitomer (2009) mentioned about the limited educational value of the accountability systems which made students to show a small part of their knowledge in artificial items which were not related with real-world contexts.

All these arguments revealed that the accountability conception is highly related with the requirements of summative assessment which focuses on 'how well students have learned' (Danielson, 2008, p. 191) in order to provide data for both decision makers and assessment users about teachers' and schools' effectiveness (Linn, 2000).

2.3.3. Student Accountability Conception

This conception consists of assigning grades for each student, comparing each student's performance with others, checking for students' progress according to the criteria determined before, selection of the students to higher level of educational institutions, informing parents or other policy-makers in the system (Brown, 2008).

However, the danger in using assessment results for high-stakes accountability purposes rather improving students' learning makes teachers and students concentrate on only aspects of competence that are likely to be assessed (William, 2001). The grading in this context did not focus on how far a student has progressed on a learning continuum but instead was merely describing where the student sat in relation to others of the same age. (Musial et al., 2009).

From another perspective, it was also said that if students were held accountable in these high-stakes tests this issue would increase the number of low achievers who gave up (Stiggins & Choppuis, 2005) and the ones who feel resentment, anxiety, lack of appropriate test-taking strategies, and decreasing motivation (Guthrie, 2002; cited in Brown, 2008).

In classrooms that teachers recorded and classified students' work and reduced it into one single score to assess students' achievement, those exam scores of students were considered as having little help in future learning of students (Weeden, Winter & Broadfoot, 2002) since students might tend to focus on what was needed to increase their scores and pass the exams (Harlen, 2007).

Such negative effects of student accountability in assessment process and therefore the reasons of why assessment users see assessment sometimes as irrelevant or pernicious will be discussed in detail in following conception, which is called "Irrelevance".

2.3.4. Irrelevance Conception

The irrelevance conception was included in the four-faceted model proposed by Brown (2002). In this model, Teacher Conceptions of Assessment, the irrelevance conception was the last one, which was discussed in detail from New Zealand teachers' perspective. In his further studies, Brown (2008) introduced detailed claims which were connected to the idea that assessment was irrelevant: (a) assessment equates testing, and by corollary, that testing is bad for education, or (b) that assessment makes teachers, schools, and students accountable for their work, which is bad for quality education (p. 25). The accountability of teachers is sometimes misinterpreted by teachers and then, they start to 'teach to test' (Berry, 2008, p. 8) in order to increase their students' scores. Berry (2008) also mentioned about the issue of high-stakes examinations, which were assumed to have negative 'backwash 'effects on teaching and learning, and saw them as a trouble because of their narrowing impact on important learning outcomes of students (p.8).

Further, the irrelevance conception might originate from the barriers that negatively affect the quality of assessment. Stiggins (1997) mentioned that "the strong negative feelings about assessment and evaluation are a barrier to quality assessment because they can prevent us from being willing to both take the risk and invest the mental energy needed to hold ourselves accountable for actual student achievement" (p.93).

Sometimes teachers' lack of enough knowledge or training about the importance of assessment techniques or how to use the assessment techniques are making the teachers consider the assessment as irrelevant. In other words, if teachers do not have enough training to implement the measurement techniques, they will find assessment of little relevance to their classroom evaluation activities (Stiggins, 1997).

The main issue in classroom environment is not that teachers are not assessing enough, but that they are not using the information they collect to help pupils learn (Weeden, Winter, & Braodfoot, 2002). In other words, little or no use of the results for modifying teaching and curriculum originated from the schools' little use of the assessment findings made teachers to file and forget about data gathered (see Gipps, Brown, McCallum, & McAlister, 1995; cited in Brown , 2008, p. 28).

2.4. The Research on Relationship between Teachers' Conceptions of Assessment and Other Variables

The studies focusing on conceptions of assessment users might be grouped as (1) teachers conceptions of assessment (Philippou & Christou, 1997; Brown, 2002; 2004; 2008; Degbey, 2009); (2) teacher trainees' conceptions of assessment (Brown, 2002; Winterbottom, Brindley, Taber, Fisher, Finney, & Riga, 2008), and (3) students' conceptions of assessment (Brown & Hirschfeld, 2008; Peterson and Irving, 2008).

To begin with; Philippou and Christou (1997) conducted a study to investigate Cypriot and Greek primary teachers' conceptions about mathematical assessment, which had some significant findings that are worth to mention in this part. Their

purposes were to have a look at teachers' opinions about "the role of assessment in mathematics learning, which grading criteria are used by teachers, what item format was the most popular to teachers, whether teachers' assessment is in line with instruction (i.e., the degree to which assessment is integral to instruction). And, their last objective was to measure the aspects of assessment in practice, which have a close relationship with their instructional objectives. In that study, both quantitative and qualitative methods were conducted to gather data from 610 (Cypriot) and 152 (Greek) primary teachers teaching fifth and sixth grade mathematics. It was mentioned that teacher were using assessment to diagnose students' learning and make decisions about the effectiveness of instruction (more improvement conception). However, most of the teachers did not conceive of assessment data as the primary source for grading purpose (less accountability conceptions), which means that they were using various sources to gather data about students' performance in class. They also examined how curriculum changes affected teachers' assessment conceptions and found that teachers were willing to follow the requirements of the curriculum; but their boundness to select activities and assessments from textbooks reveled teachers' weak understanding of assessment in instruction process.

In a larger perspective, Brown (2002) did comprehensive studies before proposing a four-faceted model for teachers' conceptions of assessment. He studied teachers' conceptions of learning, curriculum, teaching and teaching efficacy, and found out a meaningful structure that relates each of these conceptions. After studying these relations and working on validity and reliability issues of the scale that he developed,

he used Structural Equation Modeling to propose a model for teachers' conceptions of assessment. The findings in his studies revealed that three of the conceptions (Student Accountability, School Accountability, Improvement) were positively correlated, whereas the fourth major conception (Irrelevance) was negatively correlated with Improvement, uncorrelated with School Accountability, and positively correlated only with Student Accountability conception. He worked with primary level teacher trainees, undergraduate education students, practicing primary school teachers in order to have a more concrete picture for the model of teachers' conceptions of assessment, and their views were similar to each other. Moreover, the effect of assessment techniques (teacher controlled classroom assessments, formal examinations, oral assessments, and portfolio), assessment practices (deep cognitive processing, informal classroom assessment, and formal assessment), teacher roles, individual teacher characteristics (gender, role, experience, assessment training), school characteristics (size, SES), teaching level (primary and secondary) were investigated and no statistically significant mean differences among conceptions of assessment were examined.

Furthermore, Winterbottom et al. (2008) studied what assessment meant to trainee teachers who attended a Postgraduate Certificate in Education at the University of Cambridge. The d sample included 220 secondary trainee teachers taking the seminars in this university in 2007. The results of this study reported that the participants were conceiving of assessment for the purposes of (1) making learning explicit, (2) promoting learning autonomy, and (3) performance orientation. It was mentioned that the first and second factors focused on learning; however, the last one

focused on prioritizing performance gains. The study also revealed that although the trainees valued the last one as the lowest, they reported it as a stronger feature of their practices.

Degbey (2009) did a study to investigate the teachers' conceptions about the effects of different forms for assessing students' achievement outcomes. That study was a qualitative study done in Turku in Finland. There were seven teachers, specialized in Biology, Geography, Mathematics, History, Finnish, Special Education, and Literature teachers, teaching in upper secondary schools. The researcher focused on changing conceptions of assessment in Finnish education system, and gathered data to learn how teachers conceive of the assessment tools like portfolio, performance-based assessment, self and peer- assessments, observations. Teachers mentioned that these tools assessment have a good effect on students' learning, motivation, performance, and personal development. Portfolio was seen as one of the useful methods that make students select their best works and put them inside through self-assessment. Further, performance-based assessment was considered as one of the useful methods, which demands creative thinking skills, higher-order thinking skills to demonstrate, produce, develop objects which are meaningful to the students. However, when the teachers views about matriculation exam, which makes students get a score for entrance of university, was asked and teachers said that this exam affect their teaching. Teachers were highly concerned with teaching and how to adopt the assessment techniques to achieve these goals. From Finnish perspective, it was seen that the participant teachers in that study held Improvement and Accountability conceptions. As it was seen that exams in an educational system

affect the assessment users' conceptions. Since it was a qualitative study, it was hard to learn which assessment conception was common among these teachers.

When the studies about teachers' beliefs are examined in detail from Turkish perspective, it is seen that there is a limited body of research focusing on teachers' beliefs or conceptions of assessment. Specifically, the studies which are seen worth to mention in the literature might be grouped as the views of teachers and students about assessment techniques and the issues in implementations of the assessment techniques (Acar and Anıl, 2009; Bal, 2009; Erdal, 2007; Uçar, 2007); teachers' efficacy and self-efficacy beliefs towards measurement and evaluation practices (Arda, 2009; Ceylandağ, 2009; Çakan, 2004; Gelbal and Kelecioğlu, 2007); and perceptions of teachers about their application levels of measurement and evaluation (Kilmen and Çıkrıkçı- Demirtaşlı, 2009).

In Turkey, there are various studies done to investigate the relationship between teachers assessment preferences and some other variables (e.g. teachers' teaching subject, in-service training, years of teaching experience, grade level they teach, gender..etc.). Although, the Turkish researchers did not study the teacher conceptions in detail in previous years, it will be important to summarize the studies that helped researchers to find and shape the study around the variables which were included in analyses to shed light on Turkish teachers assessment conceptions.

Kaynak (2000) analyzed secondary school teachers' views about measuring and evaluating students' achievement in terms of teachers' gender, teaching area of certification, teaching experience, the type of school that they were graduated from, the type of school that they work. 223 secondary school teachers participated in that

study. The results revealed that teachers agreed on that the measurement errors in assessment process were originating from students; however, they were not sure about whether these errors were originating from themselves, the environment that assessment was done, and the assessment tool that was used in this process. Moreover, the most preferred assessment tool was essay type assessments. Also, foreign language teachers mostly preferred short-answer and true-false tests, and multiple choice tests are mostly used in Anatolian High Schools.

Pilten (2001) evaluated the elementary school teachers' understanding about measurement and evaluation practices. The data gathered through questionnaires from 211 teachers and through interviews from 35 teachers. It was found out that teachers did not get enough training about assessment techniques, and enough knowledge about how to prepare, use and apply the requirements needed for effective assessment. Further, essays, oral exams and multiple-choice tests were commonly used by elementary school teachers; and there was no significant difference among Social Studies, Science, Turkish and Mathematics teachers' assessment tool preferences.

Erdal (2007) examined the assessment dimension of the revised elementary school mathematics curriculum by investigating the in-service teachers' assessment tool preferences and their knowledge about the assessment tools suggested in curriculum. The questionnaire developed by researcher and it was conducted to 200 elementary school teachers in Afyonkarahisar in 2006-2007 academic year and interviews with four teachers were the sources to gather data for that study. Similar with Pilten's findings, the results showed that most of the in-service elementary school teachers

did not have enough knowledge and training about assessment tools suggested after curriculum revisions in 2005. Also, time limitation and lack of enough equipments in schools limited their preferences in the implementation of the assessment tools mentioned in curriculum revision process.

Uçar (2007) investigated the elementary schools teachers' views about their implementation of the assessment techniques recommended in the new mathematics curriculum. There were 306 teachers and it was seen that teachers' opinions about their implementation of the assessment techniques recommended in the new mathematics curriculum did not show a significant difference based on their teaching experiences, grade level, or class size they teach.

Bal (2009) aimed at evaluating the measurement and evaluation approaches used in fifth grade mathematics instruction in terms of the opinions of teachers and students. There were 226 primary school teachers teaching 5th grades in the center districts of Adana; and 881 fifth grade students selected randomly in these districts. Results of this study shed light on to the problems that teachers have can be stated as not having enough time, spending too much time on filling the assessment forms, lack of knowledge about assessment techniques; curriculum revisions, difficulty to work as a group, lack of filling the forms objectively, and parents' influence in doing homework. On the other side, students mentioned that they have too much project and performance workload, and have difficulty to find necessary sources and to meet as a group outside school. Further, it was revealed that observation, short-answer questions, interviews and multiple-choice questions were mostly preferred by

teachers; however, project, performance assessment, attitude scale, rubric and peer assessment were preferred less in classroom practices.

In recent decades in Turkey, teachers' self-efficacy and their self-efficacy beliefs towards measurement and evaluation were discussed in some studies. In one of the previous studies focused on this situation, Çakan (2004) compared elementary and secondary school teachers in terms of their assessment practices and their perceptions towards how well they perceive themselves in assessment process. There were 260 elementary and 244 secondary school teachers attended Educational Measurement and Evaluation Seminar in Bolu in 2003-2004 academic term. She reported that most of the teachers perceived themselves as unqualified in their measurement and evaluation practices. In addition, elementary school teachers perceived themselves more qualified than the ones work in secondary schools. She also stated that although elementary school teachers use multiple-choice items most frequently, secondary school teachers mostly prefer essay type assessment in their classrooms.

Ceylandağ (2009) contributed literature by developing an instrument to measure teacher self-efficacy toward measurement and evaluation. She worked with 394 experienced teachers working in public schools in Ankara, Samsun, and İstanbul. As the results of that study revealed that, teacher self-efficacy towards measurement and evaluation was positively correlated with frequency of using traditional and alternative measurement and evaluation tools, but years of teaching experience was a non-significant predictor for teachers' sense of efficacy, teacher self-efficacy toward measurement and evaluation, and frequency using traditional or alternative assessment tools.

To sum up, all of these studies formed a theoretical background to examine the teachers' conceptions of assessment in detail in this study in order to draw a concrete framework for how teachers conceive of assessment by referring to which assessment tools are preferred mostly.

2.5 Summary

From Turkish educational context, primary school curriculum revisions done in 2005 intended to make some changes in responsibilities of teachers and students as well. As stated by Ministry of National Education (MoNE), the idea embedded in these reforms was to have a learner-centered curriculum, which was planned to make a shift from behaviorist to a more constructivist understanding, which means that learning became more student-centered and teachers became facilitator instead of being the main source or authority in classroom environment (MEB, 2005; Erdoğan, 2007).

The shift in curriculum also made some important changes in assessment practices in classroom assessment, which focused on whole process rather than product in order to determine students' needs in learning environment (MEB, 2005). Specifically, in constructivist learning environment, the role of assessment was also revised and it gave responsibility to the teachers to use assessment to check for students' learning by monitoring students' progress and giving feedback; and the students are held responsible for their own learning through peer and self- assessment activities (Pellegrino & Hickey, 2006).

Brown's (2002) proposed model for conceptions of assessment was examined in this study and the findings related with Turkish teachers' conceptions of assessment were analyzed by taking into account the conceptions of assessment (Student Accountability, School Accountability, Improvement, and Irrelevance) included in this model.

Improvement conception mainly embedded in the idea of seeing *assessment, instruction and teaching inseparable* (Frederickson & Collins, 1989 cited in Ginsburg, Jacobs, & Lopez, 1993) and therefore making teachers carefully plan their lessons by taking into account all issues in order to improve students' learning and their teaching in assessment process (Dochy, Gijbels & Segers, 2006).

Accountability conception is highly related with the requirements of summative assessment which focuses on *how well students have learned* (Danielson, 2008) in order to provide data for both policy makers and practitioners about teachers' and schools' effectiveness (Madaus, 1985; cited in Linn, 2000). However, when assessment results were started to be used for high-stakes accountability purposes rather than improving the learning of students, there is an incentive for teachers and students to concentrate on only those aspects of competence that are likely to be assessed (William, 2001).

There were both positive and negative arguments for school and student accountability conceptions. Linn (2000) was one of the researchers who focused on positive effects of high-stakes accountability on improvement of the instruction and student learning when teachers and students are held accountable for the results of these assessments. He mentioned that assessment was one of the key components of

increasing accountability. However, Stiggins (2002) criticized about summative role of assessment which was seen useless for decision-makers who have to make decisions right, and therefore students' learning was being left out of the definition (Sangster& Overall, 2006) by the requirements of the summative rather formative assessment purposes.

Moreover, although there were various arguments to revealing assessment as irrelevance, it was seen that this conception originates from the barriers that negatively affect the quality of assessment. From teachers' perspective, Stiggins (1997) mentioned that "the strong negative feelings about assessment and evaluation are a barrier to quality assessment because they can prevent us from being willing to both take the risk and invest the mental energy needed to hold ourselves accountable for actual student achievement" (p. 93).

Finally, it was important to understand how teachers conceive of assessment, how they use the assessment techniques, and what they needed in this process by focusing on their strengths and weaknesses and support their classroom assessment practices by preparing in-service trainings or by professional development programs (Brown, 2002) in order to increase the quality of teaching and learning. Therefore, this study was conducted to examine how teachers conceive of assessment in their profession in Turkey.

CHAPTER III

METHOD

The third chapter was organized to introduce the research design, research questions, dependent and independent variables of study, subjects of the study, instrument used to gather data, data collection procedures and data analysis process, and limitations in detail.

3.1 Design of the Study

In this study, survey research was used to investigate the teachers' conceptions about the assessment by using the appropriate data-gathering tool. Since the survey studies are descriptive in nature, they are useful to investigate a variety of educational problems and issues by concerning with assessing demographics, practice, opinions, preferences, and attitudes in such related situations (Gay & Airasian, 2003).

More specifically, a cross-sectional survey served the purpose of this study to collect data in order to find out the opinions of the participants (Fraenkel& Wallen, 2006).

3.2 Research Questions

The main problem of this study was to investigate the level of teachers' conception of assessment. In order to examine how teachers conceive of assessment, the following research questions were addressed in this study:

1. What are the levels of teachers' conceptions of assessment?

2. How do the conceptions of assessment relate to each other?

3. Which assessment techniques are frequently used by teachers?

4. Is there a significant difference in teachers' conceptions of assessment according to their teaching subject (Turkish, English, Mathematics, Science and Technology, and Social Studies)?

5. Is there a significant difference in teachers' conceptions of assessment according to their years of teaching experience (5 years or less, 6-10 years, and 11 years or more)?

6. Is there a significant difference in teachers' conceptions of assessment according to in-service training?

7. Is there a significant difference in teachers' conceptions of assessment according to undergraduate institution they graduated from (The Faculty of Education and The Faculty of Arts and Sciences)?

3.3 Description of Variables

3.3.1 Independent Variables:

Years of Teaching Experience: It was one of the independent variables included to learn how many years the participants spent as a teacher. As it is known that it is continuous variable having a ratio level of measurement. In this study the value of this variable is categorized into three levels as 5 years or less, 6-10 years, and 11 years and more.

Teaching Subject: This independent variable is a categorical variable with a nominal scale for asking teachers' teaching subjects as Turkish, English, Mathematics, Science and Technology, Social Studies.

In-service Training: This independent variable is used to examine whether taking any in-service training will make any difference in teachers' conceptions of assessment. It was a categorical variable having nominal scales as yes and no.

The Undergraduate Institution: This independent variable is used to examine whether the undergraduate institution teachers graduated from make any difference in teachers' conceptions of assessment. It was a categorical variable with nominal scale with levels 'Faculty of Arts and Sciences, Faculty of Education, and Others'.

3.3.2 The Dependent Variable(s):

Teachers' Conceptions of Assessment: The main dependent variable having four subscales, Improvement, School Accountability, Student Accountability, Irrelevance, which are assessing how teachers conceive of assessment in their profession. These are all that variables having interval level of measurement. The higher mean scores for each subscale indicate that these group of teachers having higher level of each conception. The higher the mean scores for each conception revealed the higher agreement level for each conception.

3.4 Subjects of the Study

The target population of the study included the schools having more than 25 teachers in Turkish, English, Mathematics, Science and Technology, and Social Studies subject areas in public primary schools in Ankara. The schools, rather than the

participant teachers were selected through cluster random sampling from the accessible population in Ankara. The schools having at least 25 teachers teaching these subjects in spring semester of 2009-2010 academic year formed the clusters in this study. The number of teachers in these schools ranged from 25 to 37.

In order to reach the appropriate number of participants, 18 schools were visited by the researcher(s). Totally, there were 421 teachers (including 273 female and 148 males) but when outlier cases (cases 7, 23, 128, 201, 225, 389, 395) were excluded, 414 teachers were left in the sample.

As seen from Table 3.1, the English teachers (25%) had the highest percentage among the teachers. The group of teachers having five years or less experience (39%) had the highest percent among other groups of teachers. Further, majority of teachers (77%) did not take any in-service training related to assessment. In further analyses, although the groups of teachers graduated from Faculty of Education (64%) and Faculty of Arts and Sciences (27%) were included, the other group of teachers graduated from other undergraduate institutions (e.g. Teacher College, The Institute of Education) was not included because of their lower percentage (7%).

Table 3.1.

Demographic Information of Teachers (N=414)

	N	%
Teaching Subject		
English	104	25
Turkish	92	22
Mathematics	85	21
Science and Technology	71	17
Social Studies	59	14
Years of Experience		
5 years or less	161	39
6-10 years	103	25
11 years or more	145	35
In-Service Training		
Yes	97	23
No	317	77
Faculty		
Faculty of Education	264	64
Faculty of Arts and Sciences	113	27
Others	27	7

3.5 Data Collection Instrument

The data were gathered through an instrument adapted from the scale called "Teacher Conceptions of Assessment Abridged Scale" (TCoA-IIIA Abridged Scale) (see on APPENDIX B), which was the short version of Teacher Conceptions of Assessment Scale originally developed and used in English by Brown (2001-2003). There were 27 items in this abridged scale, but after pilot study, 2 items were omitted because of some reliability issues that will be explained in detail in pilot study section.

There were three main sections in the instrument given to participants in this study:

3.5.1 Demographic Information

In this section, the items about teachers' demographic information were included. There were seven questions prepared to learn the background of teachers in detail. Gender, teaching experience (levels were 1-5 years, 6-10 years, 11 years and more), teaching subject (levels were Turkish, English, Mathematics, Science and Technology, Social Studies) are asked. Moreover, whether they took any assessment course (levels were yes and no) and whether they took any in-service training about measurement and evaluation (levels were yes and no) were also examined in this section. Teachers' undergraduate institution was also asked (levels were Faculty of Arts and Sciences, Faculty of Education, and Others).

3.5.2 Measurement and Evaluation Techniques

This section included the names of the assessment techniques that were suggested to be used in curriculum revisions done by MoNE in 2005 (MEB, 2005). There were

eighteen tools (See Appendix B), and participants were made to select each one that they were using in their classroom.

3.5.3 Conceptions of Assessment Abridged Scale (CoA- IIIA Abridged Scale)

TCoA- III and TCoA- IIIA scales were developed by Brown (2001-2003, 2008) to examine the teachers' conceptions of assessment in New Zealand. The original scale had both short and long versions in English. In this study, the short version which had 27 items rated on a 6-point rating scale degreed from 1 (Strongly Disagree) to 6 (Strongly Agree) was conducted. After pilot study results, two items were excluded because, they did not work properly for the constructs that were included in the scale.

The data gathering process started by adapting the original scale by focusing on the functional/structural equivalence purposes which makes sure that the instrument measures the same psychological construct across the cultural groups (van de Vijver & Poortinga, 2005). Firstly, the original scale was translated into Turkish by three different faculty members teaching English in Ankara University, Hacettepe University. These experts had a background in measurement and evaluation by taking courses, seminars and workshops. Then, back-translation from Turkish to English was conducted by two experts in English who again teach English in Gazi University and Hacettepe University and specialized in the field of measurement and evaluation in English language teaching. In addition to these, the opinions of two different English teachers working in two public schools (in Bursa and Istanbul) and of one working in a private primary school in İstanbul were gathered in translation and back-translation steps, respectively. Furthermore, the expert opinion was taken for Turkish version of the scale and some revisions were conducted according to the

comments of these experts. There were two Turkish teaching staff in Gazi University, two curriculum and development specialists from Middle East Technical University, and three measurement and evaluation specialists working in Gazi University and Hacettepe University in Ankara.

3.5.4 Pilot Study

The pilot study was conducted in randomly selected fifteen schools whose principals permitted researcher(s) to gather data. The data were gathered in January 2010 from the teachers working in public schools in Ankara. There were 265 teachers including English (26%), Turkish (22%), Mathematics (18%), Science and Technology (17%), and Social Studies (16%) teaching subjects.

Firstly, the wording of items were checked and it is important to note that all of the items were positive statements in the tool, and therefore the reverse coding of the items was not needed before proceeding to reliability and validity analyses.

3.5.5 Reliability Analysis

The reliability analyses were conducted to check the internal consistency of the scale to find out the Cronbach's alpha coefficient. In order to better understand which items have a higher correlation with other items in the scale, item analyses process was conducted in order to check for the corrected item-total correlations. The findings showed that there were two items having corrected item-total correlations less than .3 which meant these items did not have a high correlation with the construct(s) that were included in the scale. Further, according to Cronbach's alpha if item deleted values in Appendix C, there was a significant increase when these items

which were “Formerly Item 9: Assessment results should be treated cautiously because of measurement error” and “Formerly Item 18: Teachers should account for error and imprecision in all assessment” were excluded from the scale. The reasons for this issue might originate from *item bias* (such as poor translation) or *construct bias* (such as incomplete overlap of constructs in the cultural group) (van de Vijver & Poortinga,2005), After doing this exclusion, the reliability coefficient for the scale was resulted in the value of .83, which was higher than .7 meaning a good level of consistency (Nunnally, 1978). The long version of the original TCoA-III scale (50 items) had a reliability coefficient of value of .85 in New Zealand context.

3.5.6 Validity Check

In validation process of the scale, first of all the face validity was examined by the experts, and their suggestions were highly appreciated and seriously taken into account in order to have a more understandable and trustworthy tool to gather data.

Further, 25 items of 27 items of the Teacher Conceptions of Assessment Abridged Scale (TCoA- IIIA) were subjected to Exploratory Factor Analysis (EFA) using PASW Statistics 18 (Predictive Analysis Software- Formerly SPSS). Before starting the analyses, the suitability of data and assumptions for EFA were checked.

Since the suitability of the data highly depends on the sample size, it is important to refer to the suggestions of other researchers while deciding on whether the sample size is proper to do further analyses. For instance, Tabachnick & Fidell (2007) mentioned “it is comforting to have at least 300 cases for factor analysis” (p.613), and on the other side Nunnaly (1978) stated that there might be ten cases or

participants for each item in order to properly conduct factor analysis. Taking into account these suggestions, our sample size including 265 teachers in the pilot study could be considered as a suitable sample size to move further.

In pilot study, there was a small percent of missing values, which included less than 5% of the total scores. After checking how missing values were distributed by Little's MCAR test, it was found that the missing data was randomly distributed. Based on the non-significant result (p=.90) of MCAR test, EM method was selected to deal with these random missing values.

Findings of the correlation matrix showed that all of the values of the coefficients were .3 and above, which was a good indicator of the strength of the relationship among items (Pallant, 2007). Kaiser-Meyer-Olkin (KMO) value was checked and the value of (.87), which was higher than the value of .60, was considered as a good indicator for the adequacy of the sample size for factor analysis (Pallant, 2007). In addition to KMO, it is important to report the significance (p< .05) of Barlett's Test of Sphericity , which says that there were enough evidence to conclude that correlation matrix is not an identity matrix, and therefore the correlations between each item were different than zero (Tabachnick & Fidell, 2007).

It is known that Kolmogorov-Smirnov and Shapiro-Wilk tests were conservative tests in checking normality. The findings revealing significant values *(*p<.05) made researchers to violate the assumption of normality which is assumed to be quite common in larger samples (Tabachnick & Fidell, 2007). Nevertheless, the researchers decided on checking the skewness and kurtosis values and the histograms for normality assumption. It was found that the skewness and kurtosis values of our

data were changing in a range between ±2 which means that the distribution could be assumed approximately normal (Pallant, 2007). In addition, the histograms were also checked and they were found slightly skewed which was not a serious problem to violate normality assumption.

The extreme values were checked by looking at the boxplots, although there was only one case which could be considered as a univariate outlier, but the researchers decided to keep this case during further analyses.

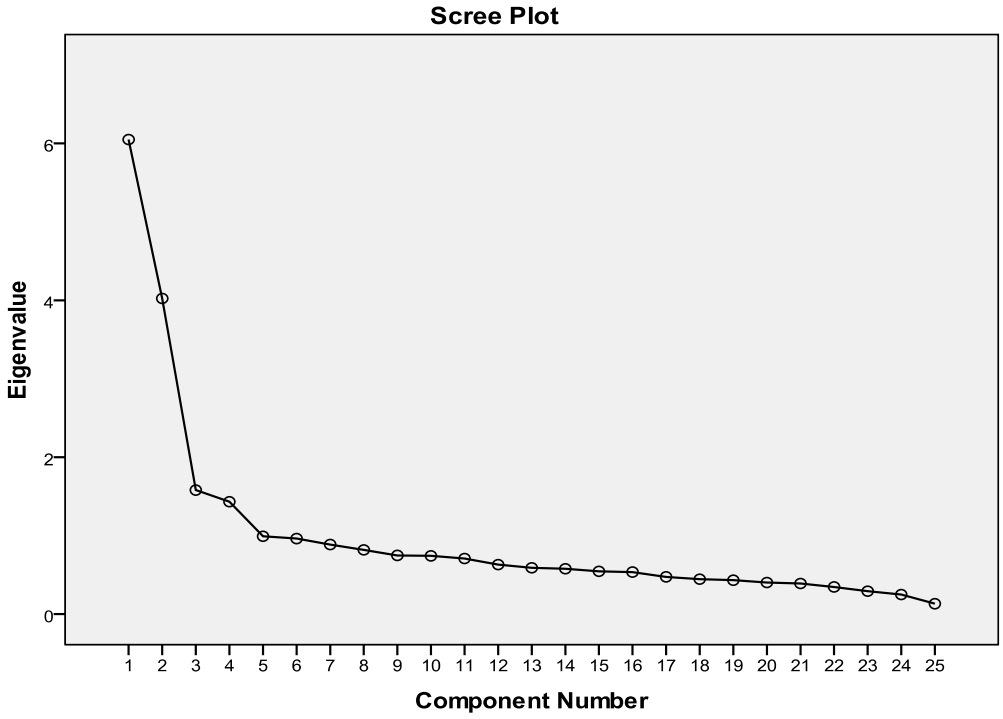
After checking all these assumptions for factor analyses, we have started to do EFA. Since we were interested in reducing the large number of items into a smaller number of components, Principal Component Analysis, a useful initial step in factor analysis, was used to extract maximum variance from the data set with each component (Tabachnick and Fidell, 2007; Thompson, 2004). Mainly, there were two rules generally taken into account in deciding to the number of factors: (1) Kaiser's criterion and (2) Catell's scree test.

According to Kaiser's criterion which makes the researchers to retain the factors with eigenvalues greater than 1 (Tabachnick & Fidell, 2007), the initial factor statistics of our data revealed that there were four factors having values higher than 1 by explaining 52.3% of the total variance. The details for the variances of the components obtained in factor analysis were given in Appendix D.

Then, scree plot our data, which makes the researchers to focus on the changes in direction of curve, (Pallant, 2007) showed that there were four eigenvalues showing

the sharp changes in the curve. Table 3.4 and Figure 3.1 show the statistics of the initial factor extraction and scree plot of our data, respectively.

Figure 3.1. Scree Plot



Moreover, since the results of factor extraction are hard to interpret, factor rotation was used to improve the interpretability of the solution (Tabachnick & Fidell, 2007). For this purpose, Varimax, which was one of the orthogonal rotation techniques, was used in order to easily group and interpret the factors. In Table 3.2, the distribution of the items after rotation can be seen clearly.

Table 3.2.

Rotated Component Matrix

Item No	Components			
	1	2	3	4
2				**,738**
10				**,758**
18	,376			**,534**
1			**,764**	
9			**,708**	
17			**,767**	
7		**,701**		
8		**,495**		
15		**,540**		
16		**,876**		
23		**,852**		
24		**,886**		
25		**,771**		
3	**,453**			,363
4	**,659**			
5	**,674**			
6	**,640**			
11	**,522**			,397
12	**,675**			
13	**,709**			
14	**,540**			
19	**,676**			
20	**,644**			
21	**,715**			
22	**,441**			,378

Although the original scale developed by Brown (2002) had both first and second order factors, the Turkish version of the TCoA- IIIA abridged scale revealed a simpler structure, which had four main factors only, therefore the issue of teachers' assessment conceptions was discussed through these main factors which were

reported in pilot study. The factors were named by their original names based on Brown's CoA-III Measurement Model of Conceptions of Assessment (2002), and New Zealand Primary teachers' Conceptions of Assessment Measurement Model-CoA-IIIA Abridged Scale (2008).

The distribution of the items with factor names can be seen in Table 3.3 which shows the reliability of each subscale and related items after pilot study. Although in the original model of Teacher Conceptions of Assessment had both first and second order factors, our data were reduced into four main factors only. This finding was different than the original model developed by Brown (2002, 2008). In the original model, Brown (2002, 2008) had four first order factors (Improvement, Irrelevance, Student Accountability, and School Accountability). In addition to those first order factors, Improvement had four second order factors (Improve Teaching, Improve Learning, Valid, Describes Ability) and Irrelevance had three second order factors (Bad for Teaching, Used but Ignored, Inaccurate) (for detail see Brown, 2002, 2008).

Table 3.3.

Reliability Coefficients of TCoA- IIIA Factors and Related Items

	Reliability	Alpha if item deleted
Student Accountability	.74	
Item 2		.70
Item 10		.73
Item 18		.69
School Accountability	.72	
Item 1		.57
Item 9		.70
Item 17		.61
Irrelevance	.87	
Item 7		.85
Item 8		.87
Item 15		.87
Item 16		.83
Item 23		.83
Item 24		.82
Item 25		.84
Improvement	.86	
Item 3		.86
Item 4		.85
Item 5		.85
Item 6		.84
Item 11		.83
Item 12		.84
Item 13		.85
Item 14		.86
Item 19		.85
Item 20		.84
Item 21		.87
Item 22		.83

3.6 Data Collection Procedure

In order to measure teacher conceptions of assessment in Turkey, the official requirements were completed successfully. Firstly, the permission from Gavin T.L. Brown, who had developed this instrument, was gathered through the permission letter that was sent via electronic mail. Then, the Turkish version of the scale was sent to the Human Subjects Ethics Committee (HSEC) and then to Educational Research and Development Head Office (ERDHO/EARGED) to get necessary research permission to gather data by using this tool. The data were gathered in beginning of May 2010 in randomly selected schools. In this process, the researchers visited each school, which was permitted. It took 10-15 minutes to respond to the items, so it was easy to collect data during teachers' short break. Also, teachers were kindly reminded to pay enough attention to respond to each item without skipping.

3.7 Data Analysis

Inferential and descriptive statistic were conducted to analyze the amount of data gathered from 421 participants. Before proceeding to further analyses, incorrect or out-of-range values, missing values and assumptions recommended by the inferential statistics were checked in order to ease the interpretation of the findings. There were a small percent of missing data which consisted of 1.7 % of the total scores in data file, and then these missing values were checked whether they form a pattern throughout the data or they were placed randomly by conducting Little's MCAR (missing completely at random) test (Tabachnick and Fidell, 2007). Although there were some other ways as ignoring the cases with missing values or deleting these cases pair-wise, it was important to keep the data in the file and so to have a larger

sample for generalizability of the results. Based on the non-significant result (p=.10) of MCAR test, EM method was selected to deal with these random missing values. As explained in Tabachnick and Fidell (2007), "E" step worked to find the conditional expectation of the missing values, and then "M" step filled these missing values with the values gathered though maximum likelihood estimation (p. 68).

In first step with overall data, Confirmatory Factor Analysis (CFA) was conducted via AMOS 18 (Analysis Moments of Structures) to check for the fit characteristics of data to the original model developed by the author who developed the original scale. Later, mean scores and standard deviations were calculated to examine the descriptive statistics for each conception of assessment. Also, frequencies were reviewed to learn which assessment techniques were commonly preferred and used by the teachers in the sample.

In order to learn the relationship between the dependent variables, which were called "Student Accountability", "School Accountability", "Improvement" and "Irrelevance"; correlation was conducted to reveal whether they have any significant correlation or not.

Further, one-way Multivariate Analysis of Variance (MANOVA) was run to answer the research questions related with teacher conceptions of assessment. Although these analyses could be done by doing a series of ANOVAs for each dependent variable, MANOVA was preferred in order to adjust for the risk for Type 1 error when there are more than one dependent variable (Pallant, 2007). Four separate MANOVAs were run to investigate the effects of teaching subject, teaching

experience, and in-service training on factors of the adapted TCoA- IIIA Abridged Scale.

3.8 Limitations

1. The study was limited with the teachers teaching Turkish, English, Mathematics, Science and Technology, and Social Studies in sixth, seventh, and eighth grades at selected primary schools in 2009-2010 academic year in Ankara.

2. The study is limited with the teachers' selected characteristics.

3. Qualitative part is not included in data gathering process in order to more deeply understand the Turkish.

CHAPTER IV

RESULTS

This chapter presents the analyses and related findings of the study by focusing on the descriptive statistics, assumption check, and inferential statistics for each dependent variable.

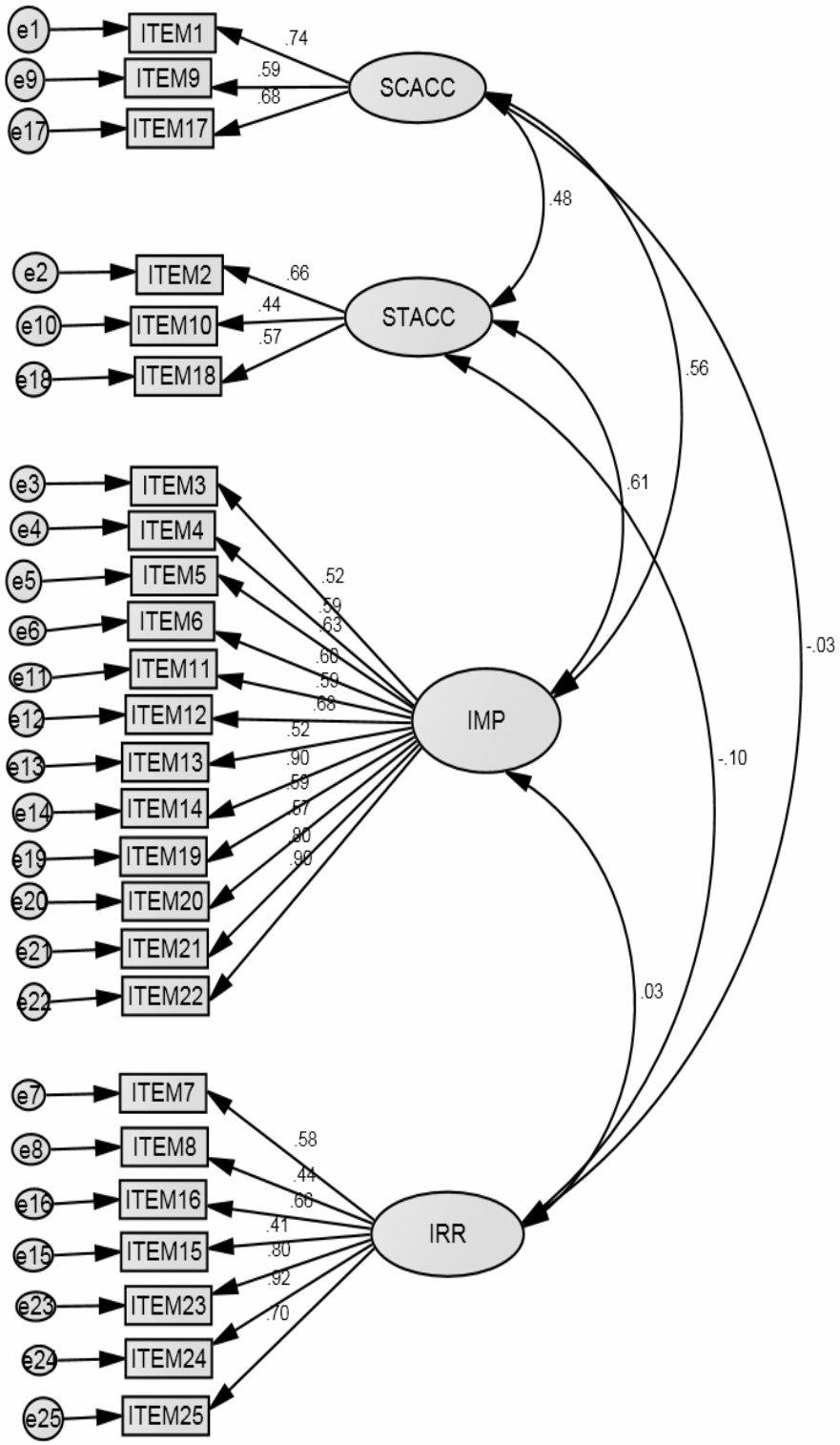
4.1. Confirmatory Factor Analysis and Reliability Analysis for Main Data

Since EFA resulted in four first order factors in this study, Confirmatory Factor Analysis (CFA) was run to check for the fit characteristics of these first order factors to the New Zealand Primary School Teachers' Conceptions of Assessment Model (both full and abridged scales) (Brown, 2002, 2008). Although the original model had both first and second order factors, EFA findings in this study revealed that sixth, seventh, and eighth grade teachers' conceptions of assessment were simpler than other teachers participated in Brown's studies in New Zealand.

In this study, CFA results reported reasonable fit characteristics with overall data (N= 414, χ^2= 695.36, df= 269, p=.00, $RMSEA$=.064, CFI=.94, $NNFI$=.93). Since RMSEA was between .05 and .80, and CFI and NNFI values were close to .95, it meant that there was a reasonable fit with these values (Thompson, 2004).

Figure 4.1 showed the standardized estimate values for the factors of the scale for data of this study.

Figure 4.1 Four Factor CFA Model of TCoA- IIIA with Standardized Estimates



The reliability analysis for overall main data resulted in similar Cronbach alpha values with pilot study. All coefficients values might be seen in Table 4.3. The reliability coefficient value for total scale was increased and resulted in .84.

Table 4.1

Reliability Coefficients of TCoA- IIIA Factors of Main Data

	Reliability	Alpha if item deleted
Student Accountability	.77	
Item 2		.75
Item 10		.72
Item 18		.66
School Accountability	.73	
Item 1		.64
Item 9		.70
Item 17		.71
Irrelevance	.87	
Item 7		.82
Item 8		.86
Item 15		.86
Item 16		.83
Item 23		.78
Item 24		.80
Item 25		.84
Improvement	.88	
Item 3		.82
Item 4		.79
Item 5		.81
Item 6		.84
Item 11		.73
Item 12		.78
Item 13		.84
Item 14		.81
Item 19		.85
Item 20		.81
Item 21		.76
Item 22		.77

4.2. Descriptive Results for Conceptions of Assessment

In this part, the main question for determining the kind of conceptions teachers have about assessment was investigated. In the Table 4.1., descriptive statistics for the agreement level of teachers for each component of the Teacher Conceptions of Assessment Abridged Scale (TCoA-IIIA) was given. For this scale, the minimum value was 1, and the maximum value for each response was 6.

Table 4.2

Agreement Level of Teachers for Components of TCoA- IIIA Scale (N=414)

Dependent Variables		*M*	*SD*
Student Accountability	(STACC)	4.71	.86
School Accountability	(SCACC)	4.11	1.07
Improvement	(IMP)	4.70	.71
Irrelevance	(IRR)	3.88	1.31

As shown in Table 4.2, there were four conceptions of assessment included in the scale. Student Accountability (*M*=4.71, *SD*=.86) and Improvement (*M*= 4.70, *SD*= .71) had the highest two agreement levels which were considered as being around "Moderate Agreement" level among other variables. The Irrelevance conception (*M*= 3.88, *SD*= 1.31) had the smallest mean which could be considered as being around "Slight Agreement" level among other variables.

4.3 Results for Frequency of Assessment Tools

The research question about which assessment tools were preferred mostly by teachers teaching Turkish, English, Mathematics, Social Studies, Science and Technology to 6^{th}, 7^{th}, and 8^{th} grades in primary schools was answered in this part.

As an indicator for teachers' conceptions of assessment, frequencies were calculated to examine which assessment tools were commonly used in assessment process.

Table 4.3

Frequency of Teachers' Assessment Tools Preferences (N=414)

Assessment Techniques	*N*	*%*
Multiple Choice	402	97
Performance Task	361	87
Fill-in Blank	361	87
True-False	358	87
Project	353	85
Short Answer	327	79
Matching	319	77
Group Work	320	77
Presentation	296	72
Observation Form	273	66
Portfolio	247	60
Long Answer	230	56
Drama	215	52
Self-Assessment	184	44
Peer- Assessment	176	43
Rubric	67	16
Oral Exam	30	7
Constructed Grid	16	4
Other	18	4

As shown in Table 4.3, Multiple Choice (97%), Performance-Task (87%), Fill-in-the Blanks (87%), True-False (87%), Project (85%), Short Answer (79%), Group-Work

(77%), Matching (77%), Portfolio (60%), and Drama (52%) were the assessment tools commonly preferred by teachers teaching these subjects.

However, Constructed Grid (4%), Oral-Exam (7%) and Rubric (16%) were the less preferred assessment techniques.

4.4. Inferential Results

4.4.1. Correlation Results

In order to investigate the relationship between the measures of teacher conceptions of assessment which are Student Accountability (STACC), School Accountability (SCACC), Improvement (IMP), and Irrelevance (IRR); Pearson product-moment correlation coefficient was used.

Before running the correlations, it was important to do the preliminary analyses to check for the suitability of the data. To begin with, all of the responses given to the items were based on the Likert type responses (from 1-Strongly Disagree- to 6-Strongly Agree), and the level of measurement of the factors were all interval. In addition, these variables were all metric, and therefore these four factors were obtained by simply averaging the each participant's responses given for each factor. Further, it was also assumed that each subject provided a score for each variable which satisfied the condition to have scores for each variable belong to the same participant.

In data gathering process, since the schools included in the sample were visited individually by the researcher(s), the teachers were observed independently while they were filling the questionnaires during their breaks. This procedure was also

helpful to have lower number of missing values since the researcher(s) immediately checked for the missing values in the questionnaire and then kindly reminded them to fill in the items they skipped or did not see. Although the researchers paid a great effort to decrease the number missing values, unfortunately there were some missing values. Distribution of the missing values were checked through Little's MCAR test to be sure whether they were random or following a pattern. EM missing value analysis method was conducted to fill in the missing values with an estimated value.

The descriptive statistics helped to find and to deal with the extreme cases which are called univariate outliers (Tabachnick & Fidell, 2007). Standardized z-scores (range between ±3.29) and graphical representations were checked in detecting these outliers in data file. In Student Accountability factor, there were two cases with extremely low z-scores having values of-3.54 and -3.89; and three cases in Improvement factor with extremely low z scores which are -3.59, – 3.70, and -3.71. Although the cases were checked one more time by comparing data file and questionnaires to be sure whether there are any incorrect entries or not, the researchers decided to delete these five out-of-range cases.

After deleting these cases, the multivariate outliers were checked through Mahalanobis distance. As explained in Pallant (2007), the column produced for checking this distance was compared with the critical value, which was 18.47, and the cases having greater values than this critical score were considered as to be multivariate outliers which negatively affect the normality of the data. The analyses reveled that there were two cases having values of 18.92 and 23.69 which are exceeding the critical value mentioned above. Then, these two values were excluded

from the data file. Totally, these screening attempts made researchers delete seven cases, which resulted in remaining 414 cases, and move on doing other analyses with a well distributed data file.

Moreover, it is best to discuss the findings related with normality assumption. Kolmogorov-Smirnov and Shapiro-Wilk tests were examined; and unfortunately, these significant results were the indicators of non-normal distribution of the data. Since these strict tests were not the only tools to check for normality, the researchers focused on finding some other evidences to meet this assumption. The histograms for each dependent variable with normal curves were checked and it was seen that each distribution was almost normal with a slightly skewed direction. More specifically, the values showing skewness, related with how symmetrical the distribution is, changed from -.23 to -.75; and values showing the kurtosis, related with peakedness of the distribution, changed from -1.20 to +.36. All of the Skewness and Kurtosis values were given in Table 4.4.

Table 4.4

Skewness and Kurtosis Values for Each Variable

	Skewness	Kurtosis
STACC	-.75	36
SCACC	-.49	-.22
IRR	-.23	-1.20
IMP	-.55	.20

In order to check for the linearity assumption, scatter plots were checked and an oval-shaped or elliptical scatter plot was seen for each pairs of the variables.

The assumption for homoscedasticity, which means that the variability in scores for one continuous variable is roughly the same at all values of another continuous variable (Tabachnick &Fidell, 2007), is again checked by looking at the scatter plots. This assumption was not violated since each plot had similar shape which was oval or elliptical.

As stated in each step of the preliminary analyses, none of the assumptions of normality, linearity, and homoscedasticity were violated. In Table 4.5, correlations for each dependent variable were given. It is important to report that the Irrelevance conception of assessment had non-significant relationship with other three variables. On the other side, all other pairs of the variables were moderately correlated with each other. Since these dependent variables were the components or factors of the same scale, MANOVA test was run for each of the dependent variables including Irrelevance conception.

Table 4.5

Correlations among Teacher Conceptions of Assessment (N=414)

	1	2	3	4
STACC	1	.30**	.40**	-.07
SCACC		1	.45**	-.02
IMP			1	.04
IRR				1

** $p < 0.01$ (2-tailed).

4.4.2. Multivariate Analysis of Variance (MANOVA) Results

Most of the preliminary analyses or assumptions (Sample Size, Missing Data, Independence of Observation, Linearity, Normality, Outliers) which were checked and explained in Correlation section were also the requirements of Multivariate Analyses of Variance (MANOVA). Hence, in order to prevent the repetition of the explanations of same assumptions, they will be explained only once.

Moreover, the findings for the assumptions of homogeneity of variance-covariance matrices will be explained in following research question separately.

4.4.2.1. Results related with Teaching Subject

The question asking for whether there is a significant difference between teachers' conceptions of assessment according to their teaching subject (Turkish, English,

Science and Technology, Mathematics, and Social Studies was investigated in this part.

The homogeneity of variance-covariance matrix assumption for the third question was checked through Box's M and Levene's test. The non-significant value of Box' M, $F(40, 27842.6)= .72$, $p> .05$, was the indicator of testing the hypothesis that the variance –covariance matrices were the same in the groups, and therefore it revealed that the assumption of homogeneity of variance-covariance matrices was met (Field, 2005). However, as it was seen in Table 4.6, the results of Levene's test for each dependent variable revealed a significant result for Irrelevance conception only.

Table 4.6

Levene's Test Results for Teaching Subject

	F	*df1*	*df2*	*p*
STACC	.615	4	406	.65
SCACC	.950	4	406	.44
IRR	2.659	4	406	.03
IMP	.764	4	406	.55

Although there were some differences in mean scores (see Table 4.7) of the conceptions with respect to the teachers' teaching subject, these differences in descriptive results were not considered to be significant values by the findings of MANOVA.

Table 4.7

Descriptive Statistics for Teaching Subject

		STACC		SCACC		IRR		IMP	
Teaching Subject	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
English	104	4.64	.89	4.13	1.08	3.97	1.25	4.68	.70
Sci. and Tech.	71	4.64	.94	4.25	1.05	3.99	1.20	4.66	.80
Mathematics	85	4.75	.81	3.99	1.14	4.07	1.29	4.66	.69
Social Studies	59	4.80	.86	3.92	1.13	3.57	1.48	4.65	.69
Turkish	92	4.74	.83	4.18	.94	3.77	1.32	4.80	.67

The results revealed that teachers' conceptions about assessment did not change according to their teaching subjects which were Turkish, English, Mathematics, Science and Technology, and Social Studies. Although there were not any significant changes in descriptive statistics of these teachers' conceptions, their mean values had some differences for each group of teachers. For instance, although Social Studies (*M*= 4.80, *SD*= .86) and Mathematics (*M*= 4.75, *SD*= .81) teachers had the highest means values in Student Accountability conception; and English (*M*= 4.64, *SD*= .89) and Science and Technology (*M*= 4.64, *SD*=.94) teachers had the lowest mean values in the same conception. Then, for School Accountability conception, Turkish (*M*= 4.18, *SD*= .94) and Science and Technology (*M*=4.25, *SD*= 1.05) teachers had the highest; and Social Studies (*M*= 3.92, *SD*=1.13) and Mathematics (*M*= 3.99, *SD*= 1.14) had the lowest mean scores. In Improvement conception, Turkish (*M*= 4.80, *SD*=.67) and English (*M*= 4.68, *SD*= .70) teachers had the highest mean scores; and

Science and Technology (*M*= 4.66, *SD*=.80), Mathematics (*M*=4.66, *SD*= .69), Social Studies (*M*=4.65, *SD*= .69) teachers had the lowest mean values.

Table 4.8

MANOVA Results for Teaching Subject

Effect	*Wilk's Lambda*	*F*	*Hypothesized df*	*Error df*	*p*	η^2
Teaching Subject	.32	1.128	16.000	1624.000	.32	.01

Finally, MANOVA results revealed that teachers' teaching subject did not have a significant effect on teachers' conceptions of assessment.

4.4.2.2 Results related with In-Service Training

The question asking for whether there is a significant difference between teachers' conceptions of assessment according to in-service training was examined in this part.

In order to satisfy the assumption of homogeneity of variance-covariance, again Box's M and Levene's tests were checked. The Box's M was non-significant with the value of F (10, 45822.781) =1.328, p> .05. In addition, the non-significant results of Levene's test results were shown in Table 4.9.

Table 4.9

Levene's Test Results for In-Service Training

	F	*df1*	*df2*	*p*
STACC	1.390	1	412	.24
SCACC	.069	1	412	.79
IRR	.570	1	412	.45
IMP	1.204	1	412	.27

After satisfying the requirements of assumption check carefully, the findings of MANOVA again revealed that in-service training had a non-significant effect on the dependent variables included in Teacher Conceptions of Assessment Abridged Scale. The Table 4.10 summarized the main findings of MANOVA related to in-service training.

Table 4.10

MANOVA Results for In-Service Training

Effect	*Wilk's Lambda*	*F*	*Hypothesized df*	*Error df*	*p*	η^2
In-Service	.995	.470	4.000	409.000	.76	.01

When the mean scores of groups were checked, it was seen that each mean scores for the dependent variables were close to each other as shown in Table 4.11.

Table 4.11

Descriptive Statistics for In-Service Training

In-Service		STACC		SCACC		IRR		IMP	
Training	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Yes	97	4.72	.96	4.10	1.1	3.95	1.30	4.76	.75
No	317	4.70	.83	4.12	1.06	3.86	1.32	4.68	.69

4.4.2.3.. Results related with Years of Teaching Experience

The question asking for whether there is a significant difference between teachers' conceptions of assessment according to the teachers' years of experience (5 years or less, 6-10 years, and 11 years or more was examined in this part.

Box's M and Levene's test were examined to check for the homogeneity of variance-covariance assumption. It was seen that Box's M value had a significant value as F (20, 44518.008) = 3.315, p<.05, which had to be non-significant in order to satisfy the condition to meet the assumption of homogeneity of variance-covariance.

Table 4.12

Levene's Test Results for Years of Teaching Experience

	F	*df1*	*df2*	*p*
STACC	2.996	2	406	.05
SCACC	2.055	2	406	.13
IMP	2.974	2	406	.05
IRR	9.029	2	406	.06

Apart from Box' M, Levene's test results were checked in Table 4.12, and each non-significant value made sure the researchers meet the assumption for homogeneity of variance/ covariance matrix.

Table 4.13

Descriptive Statistics for Years of Teaching Experience

		STACC		SCACC		IRR		IMP	
Experience	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
5 years or less	161	4.53	.93	3.99	1.08	4.34	1.09	4.64	1.09
6-10 years	103	4.76	.81	4.16	.93	3.88	1.34	4.75	.60
11 years or more	145	4.85	.79	4.20	1.14	3.40	1.33	4.70	.76

The main descriptive results of these variables were given in Table 4.13, and according to this table there were some mean score differences between conceptions

of assessment. In order to check for the significance for this mean difference, MANOVA test was run.

Table 4.14

MANOVA Results for Years of Teaching Experience

Effect	*Pillai's Trace*	*F*	*Hypothesized df*	*Error df*	*p*	η^2
Years of Experience	.124	6853	8000	808.000	.00	.06

MANOVA made a significant difference with Pillai's Trace, having the value of .877. The reason for choosing Pillai's Trace is that it is becoming robust than the other three statistics (Wilk's Lambda, Hotelling' Trace, and Roy's Largest Root) when separation of groups is distributed over dimensions and the assumption of homogeneity of variance-covariance matrices was violated (Tabachnick & Fidell, 2007).

The significant value of the Pillai's Trace statistics made researchers to check for the follow-up analyses (Field, 2005, p. 606) through Bonferroni by dividing the alpha level to the number of the dependent variables in order to control for the Type 1 error (Green, Salkind, and Akey, 2000). Based on this new alpha value (= .05÷4= .0125), the univariate ANOVA results were checked and it was seen that two of the dependent variables were significant. The Table 4.15 revealed the details for each variable.

Table 4.15

Tests of Between-Subjects for Years of Teaching Experience

Source	*Variable*	*df*	*F*	*p*	η^2
Years of	STACC	2	5,918	.00	.03
Teaching	SCACC	2	1,648	.19	.01
Experience	IMP	2	.768	.46	.00
	IRR	2	22.047	.00	.10

The results showed that the years of teaching experience made a significant difference in Student Accountability and Irrelevance conceptions with significance value of .00 for those conceptions.

When Bonferroni multiple comparison results were checked for the groups of teachers having "5 years or less", "6- 10 years" and "11 years or more" teaching experience; it was seen that teachers having 11 years or more teaching experience (*M*= 4.85, *SD*= .79) had significantly highest level of "Student Accountability" conception in comparison with the other two groups of teachers. However, the other groups or teachers were not significantly different from each other.

Further, multiple comparisons for Irrelevance conception revealed significant differences among each group of teachers with respect to their years of teaching experience. The mean score of group of teachers having 5 years or less teaching experience (*M*= 4.34, *SD*=1.09) had significantly highest level of Irrelevance conception in comparison with other two groups. In addition, the group of teachers

having 6-10 years teaching experience (*M*=3.88, *SD*= 1.34) had significantly higher level of "Irrelevance" conception than the group of teachers having 11 years and more teaching experience (*M*= 3.40, *SD*= 1.33).

4.4.2.4 Results related with Undergraduate Institution Teachers Graduated from

The question investigating whether there was a significant difference between teachers' conceptions of assessment according to the undergraduate institution that teachers graduated from (Faculty of Education and Faculty of Arts and Sciences) was examined in this part. Since there was a small number of teachers graduated from other institutions (e.g. Institute of Education), this group was not included in following analyses.

Box's M and Levene's test were examined to check for the homogeneity of variance-covariance assumption. It was seen that Box's M value had a non-significant value as F (10, 201666.698) = 1.816, p>.05. Later, as shown in Table 4.14., Levene's test results were checked and each value was non-significant for each dependent variable, therefore the researchers decided to move to the other steps of MANOVA.

Table 4.16

Levene's Test Results for Undergraduate Institution

	F	*df1*	*df2*	*p*
STACC	4.525	1	357	.06
SCACC	.422	1	357	.52
IMP	.397	1	357	.53
IRR	10.342	1	357	.07

Table 4.17

MANOVA Results for Undergraduate Institution

Effect	*Wilk's Lambda*	*F*	*Hypothesized df*	*Error df*	*p*	η^2
Undergraduate Institution	.96	3.889	4.000	354.000	.00	.04

MANOVA results revealed that there is significant difference between these two groups of teachers graduated from Faculty of Arts and Sciences and Faculty of Education.

Table 4.18.

Tests of Between-Subjects Effects for Undergraduate Institution

Source	*Variable*	*df*	*F*	*p*	η^2
Undergraduate	STACC	1	6.388	.01	.02
Institution	SCACC	1	2.035	.16	.01
	IMP	1	3.795	.05	.01
	IRR	1	7.940	.01	.02

The results showed that the institution that teachers were graduated from made a significant difference in Student Accountability and Irrelevance conceptions with significance value of .01 for these conceptions.

Table 4.19

Descriptive Statistics for Undergraduate Institution

	STACC		SCACC		IRR		IMP	
Undergrad. Ins.	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Fac. of Education	4.60	.90	4.07	1.07	4.07	1.26	4.67	.68
Fac. of Arts and Sciences	4.84	.76	4.21	.1.04	3.67	1.43	4.80	.66

The teachers graduated from "Faculty of Arts and Sciences" (*M*= 4.84, *SD*= .76) had significantly higher level of "Student Accountability" conception than the teachers from Faculty of Education (*M*=4.60, *SD*=.90).

Further, the group of teachers graduated from "Faculty of Education" (*M*= 4.07, *SD*=1.26) had significantly higher level of "Irrelevance" conception than the ones came from "Faculty of Arts and Sciences" (*M*= 3.67, *SD*= 1.43).

4.5 Summary

In order to answer the research question included in this study, both descriptive and inferential statistical analyses were conducted. Mainly, descriptive findings were analyzed through frequencies, mean, and standard deviations. There were four main conception of assessment (Student Accountability, School Accountability, Improvement, and Irrelevance) having different mean scores and standard deviations. Student accountability had the highest, and Irrelevance had the lowest mean scores among other conceptions. Then, the relation between these conceptions of assessment was examined by correlation. It was found that Student Accountability, School Accountability and Improvement conceptions were moderately correlated with each other. However, Irrelevance conception did not have a significant relationship with other conceptions.

When the frequencies to check for which assessment tools were examined, it was found that Multiple Choice, Performance-Task, Fill-in-the-Blank, True-False, Project, Short Answer, Group-Work, Matching, Portfolio and Drama were the assessment tools commonly preferred by teachers teaching these subjects, respectively. However, Constructed Grid, Oral-Exam, and Rubric were the ones least preferred by teachers.

The findings of Multivariate Analysis of Variance (MANOVA) indicated that teaching subject and in-service training did not make any significant differences between teachers' conceptions of assessment. However, years of teaching experience did a significant difference in Student Accountability and Irrelevance conceptions. According to Bonferroni test, teachers having 11 years or more experience had the highest mean score for Student Accountability and lowest mean score for Irrelevance conceptions. Moreover, teachers with five years or less experience had the highest mean score for Irrelevance conception. There were also some significant differences between groups of teachers graduated from Faculty of Education and Faculty of Arts and Sciences in Student Accountability and Irrelevance conceptions.

The teachers graduated from Faculty of Arts and Sciences had significantly higher level of Student Accountability conception than others; whereas the group of teachers graduated from Faculty of Education had significantly higher level of Irrelevance conception than the ones came from Faculty of Arts and Sciences.

CHAPTER V

DISCUSSION

In this part, all findings were discussed with respect to the research questions and findings in this study. Based on the findings and limitations, implications for practice and recommendations for further studies were included.

5.1. Discussion of Descriptive Results and Correlation of Conceptions

The descriptive results of each dependent variable in Teacher Conceptions of Assessment Abridged Scale revealed that Student Accountability conception had the highest (*M*=4.71, *SD*=.86), whereas Irrelevance conception had the lowest mean (*M*=3.88, *SD*= 1.31) scores among other conceptions. The highest score for Student Accountability conception of teachers might originate from the competitive nature of Turkish education system that makes students spend higher effort to have higher grades at school and higher scores in nationwide examinations. Since these high-stakes assessments have an enormous effect on shaping the future plans of students, the teachers also take this responsibility seriously to prepare their students to these exams and, therefore make their students get higher scores than others. Currently; at the end of each academic year, 6^{th}, 7^{th} and 8^{th} grade students are taking an exam which is called Student Level Determination Exam (LDE/SBS). In these exams for each grade level, the students were only given their scores without any comparison with other students. Although these implications were planned to decrease the competitive nature of Turkish assessment system, the attitudes of students, teachers and parents towards these assessments did not change. For instance, parents are still

paying for private institutions, courses or tutors to increase their children's scores in these exams or teachers are still using various classroom examinations to compare their students. The reason of such implications lies behind the ideas of (1) mismatch between the intended curriculum and the score of assessment (Harlen, 2008; Sağlam-Arslan et al., 2009), and (2) summative assessments' narrowing effect on the curriculum and teaching strategies (Harlen, 2008). This issue was also considered as one of the dangers in education which unfortunately makes teachers teach to the test (Black and William, 1998) by teaching their students how to pass the tests with their surface knowledge rather improving their deeper understanding (Berry, 2008, p. 8).

Moreover, correlation results revealed that Student Accountability conception was moderately correlated with both Improvement *(r*=.40, *p*<.05 at α=.01), and School Accountability (*r*=.30, *p*<.05 at α=.01) conceptions. From this perspective, it can be said that teachers in this group conceived of assessment as assigning a grade or placing students into categories in order to increase their students' scores in assessments. TIMMS (1999) findings for our country might be considered parallel with this study. In this report, it was mentioned that most of teachers in Turkey were using assessment to assigning grades to students, whereas they were rarely using assessment to give feedback to students and families (cited in Kilmen and Çıkrıkçı, 2009).

The Improvement conception had the second higher value (*M*= 4.70, *SD*=.71), and this close but lower mean value of this conception made researchers to look at and discuss the situation much more carefully from a more optimist perspective. By taking into account the curriculum revision efforts after 2005, it might be said that

teachers were learning how effectively to use assessment to improve students' learning and their teaching in their classrooms but still they need time to change their Student Accountability conceptions by practicing various assessment techniques which focus on increasing students' learning. However, this change requires time to make teachers internalize these efforts, which were new and unfamiliar to them.

The correlation results for Irrelevance conception reported that this conception of assessment did not have any significant relationship with other conceptions. Moreover, the lowest mean score for Irrelevance conception (*M*= 3.88, *SD*= 1.31) among other variables was showing that assessment was not seen as irrelevant in teaching and learning environment by these teachers. When the results of Brown's (2002, 2008) studies were compared, it was seen that there were some differences in correlations of these conceptions. Brown (2002, 2008) reported that Irrelevance conception had a non-significant correlation with School Accountability, a significant negative correlation with Improvement (*r*=-.75), and moderate correlation with Student Accountability conception (*r*=.30). Further, there was also a correlation between School and Student accountability factors (*r*=.60). Then, Improvement had a positive correlation with School Accountability (about *r*=.60), and with Student Accountability (about= .30) conceptions. Based on his findings, the participant teachers in his studies were orienting assessment for improvement with improving school quality rather assigning grades to students.

Moreover, Philippou and Christou (1997) investigated Greek and Cypriot primary teachers' conceptions about mathematics assessment and they found that those teachers in the study conceived of the role of assessment as to (1) diagnose students'

learning and then (2) make decisions about the effectiveness of instruction, which were the indicators of improvement conception of assessment among these groups of teachers. However, these teachers did not conceive of assessment as the main source for assigning grades to students. When these findings were compared with Turkey, the results of these two studies were different than Turkish teachers' conceptions of assessment.

5.2. Discussion with respect to Teaching Subject

There were differences in mean scores of each conception but these differences were not significant according to the findings gathered in analysis process.

The reason for this issue might be explained as; each group of the teachers in this group were teaching the subjects that were included in the national exams of 6th, 7th, and 8th grades, and therefore each group of the teachers have had similar pressure to increase their students' scores in these exams. However, if the study was conducted with the teachers teaching 1st- 5th grades and the ones teaching 6th- to 8th grades, there might be differences in their conceptions with respect to their teaching subject. Although it was not directly related with teachers' conceptions of assessment, it might be helpful to mention about the findings of İşler's (2008) study to make concrete how those national examinations affect teachers' efficacy beliefs with respect to their area of teaching. She found out that primary schools teachers were much more efficacious than the mathematics teachers teaching 6th, 7th, and 8th grades. Then, she discussed national examinations and their negative effects on efficacy beliefs of teachers. She said that mathematics teachers focused more on the scope of the exams in classroom and paid less attention to the requirements of new curriculum

(p. 91). From this perspective, it might be said that since assessment is an important component of curriculum, the contents of both curriculum and assessment should be parallel to increase teachers' efficacy beliefs in teaching and assessment, which will increase positive conceptions towards assessment.

5.3. Discussion with respect to In-Service Training

The results revealed that conceptions of teachers about assessment did not change with respect to the in-service training they took on measurement and evaluation. In this study, in-service training about assessment was not categorized into smaller parts. But in some studies, the assessment literacy training was considered as a general term and the effect of assessment training was checked by taking into account the separate effect of all these trainings. For instance, in some studies done abroad, the effect of the amount of assessment type was examined by categorizing the amount as "Some hours in pre-service training, half to one day workshop or seminar, completed undergraduate paper, completed postgraduate paper", and no difference was found among these amounts (Brown, 2002, 2008).

In this study, there were 97 teachers who took in-service training and 317 teachers who did not get any training in assessment. This inconsistency was revealing the truth about this issue because there was a high number of teachers who did not get such training about assessment techniques and who really needed to be informed about the assessment tools recommended after curriculum revisions.

This result was consistent with the findings of Brown (2002), who mentioned about such trainings did not make any differences in teachers' conceptions of assessment.

On the other side, it might be worth to discuss the content, quality and effectiveness of these trainings. It was said that these trainings about assessment (1) included lots of theoretical rather practical knowledge and (2) done in crowded groups might be a reason for the lower quality of these trainings (Çelikkaya, Karakuş, & Demirtaş, 2010).

On the other side, there were studies discussed effects of these trainings on the frequency of assessment tools that teachers prefer to use. Çelikkaya, Karakuş, and Demirtaş (2010) mentioned that there were not any significant differences in mean scores of the teachers' preferences in use of assessment tools, suggested during curriculum revisions, with respect to in-service training. When compared with the teachers who had assessment training with the ones who did not, Social Studies teachers having assessment training were more frequently using presentations, portfolios, observation forms, peer- assessment and group work than other group with no training. They also said that in-service training did not change the frequency of using the multiple-choice, performance-based assessment, project, self-assessment, and interviews in assessment process.

5.4. Discussion with respect to Years of Teaching Experience

The results revealed that years of teaching experience (5 years or less, 6-10 years, and 11 years or more) made a significant difference on only two conceptions of assessment. There were some differences in Student Accountability and Irrelevance conception mean scores of the teachers in these groups. Based on the findings of Bonferroni test results, the more experienced group of teachers had the highest level of Student Accountability conception than the other less experienced groups.

Although curriculum revisions in Turkey aimed at enlarging the idea of assessing teaching and learning process in a student- centered learning environment rather than the product itself, the closer mean scores of teachers' student accountability and improvement conceptions revealed that the idea behind these revisions which took its roots from constructivist approach was getting enlarged among teachers' practices and conceptions. The reason of high mean scores of the most expert teachers' student and school accountability conceptions could originate from their previous experiences, which might be based on more product-based assessment strategies to increase their students' scores in exams.

However, when the mean scores for Irrelevance conception were checked, it was seen that the teachers having five years or less teaching experience had the highest mean value for irrelevance conception. This result was surprising and might be considered as the results of less quality of assessment courses in universities.

When this finding was compared with Brown' study (2002), it was seen that years of teaching experience of teachers did not revealed a statistically significant difference for teacher conceptions of assessment.

In Turkey, Arda (2009) examined efficacy beliefs and views of elementary schools teachers related view the assessment dimension of 2005 mathematics curriculum, and found out that the most expert teachers, who have an average of 14.20 years of teaching experience, were more sensitive than the younger teachers, with a 9.26 year-teaching experience, to follow and examine the changes in curriculum after 2005 curriculum revisions. This result might be an indicator of these teachers' higher motivation for teaching, otherwise it might be the opposite to have more experienced

teachers with teachers who was less open to follow changes and less motivated to use new teaching and assessment strategies in their classrooms.

5.5. Discussion with respect to Undergraduate Institution Teachers Graduated from

The differences in Student Accountability and Irrelevance conceptions were significant for the groups of teachers coming from Faculty of Education and Faculty of Arts and Sciences. For Student Accountability conception, the teachers graduated from Faculty of Arts and Sciences (*M*= 4.86, *SD*=.76) had higher mean scores than the teachers graduated from Faculty of Education (*M*=4.60, *SD*=.90).

For Irrelevance conception, the teachers from Faculty of Education (*M*= 4.09, *SD*=1.25) had higher mean scores than the teachers graduated from Faculty of Arts and Sciences (*M*= 3.66, *SD*= 1.44).

It might be said that the training given in a small period of time to those teachers graduated from Faculty of Arts and Science to become teachers might be the reason for having high level of Student Accountability conception, since this short period might not be enough to make these teachers internalize the significance of using assessment in overall process rather evaluating the product. Özdemir (2009) mentioned that the teachers came from other faculties than Faculty of Education had negative attitudes towards teaching and teaching profession.

5.6. Discussion of Assessment Preferences of Teachers

Learning which assessment tools were mostly preferred by teachers was a good indicator about learning teachers' conceptions of assessment. Frequencies were

calculated to examine which assessment tools were commonly used in assessment process.

Statistics revealed that Multiple Choice (97%), Performance-Task (87%), Fill-in-the Blanks (87%), True-False (87%), Project (85%), Short Answer (79%), Group-Work (77%), Matching (77%), Portfolio (60%), and Drama (52%) were the assessment tools commonly preferred by teachers teaching these subjects. When these results were compared with other studies done before and after 2005 curriculum revisions in Turkey, the commonly used tools did not change a lot. Since it was known that teachers had tendency to use the tools especially which they knew or had higher efficacy beliefs towards using them (Gelbal & Kelecioğlu, 2007), most of the participant teachers in this study selected the traditional assessment tools except for performance-based assessment, project, and portfolios.

However, Constructed Grid (4%), Oral-Exam (7%) and Rubric (16%) were the less preferred assessment techniques. Although teachers needed to use rubric while they were assessing performance-based assessments, projects or portfolios, the lower percent of using rubrics was contradictory with this finding. The reasons for this issue might be due to that teachers did not have enough knowledge about how to use rubrics or how to grade students' performance assessments or projects when they gave such assessment tasks to their students.

When the findings of Brown (2002) were examined for the effect of assessment techniques on teachers' conceptions of assessment, he reported statistically non-significant effects of assessment techniques on teachers' conceptions of assessment.

5.7. Implications for Practice

Based on the findings of this study, novice teachers having five years or less teaching experience had the highest mean scores in Irrelevance conception, therefore the experts in MoNE and instructors teaching in universities need to work hand in hand to help these teachers. Moreover, the quality of assessment courses in university programs should be increased in order help all teacher candidates to see the importance of assessment in teaching and learning environment. Apart from the revisions in theoretical background of these courses, the practice side should be improved in order make the university students learn how to use the assessment tools which are new to them before going to the real classrooms.

In order to make teachers adapt much more easily to use the assessment tools recommended in curriculum revision process in 2005, the quality and content of in-service trainings, seminars or workshops related with assessment should be increased. Since teacher change takes time, teachers should be given opportunities to face their strengths and weaknesses in assessment process and become much more familiar with how they conceive of assessment in their profession.

5.8. Recommendation for Further Research

This study was limited with the sample selected from Ankara only. In further studies, more teachers from other cities of Turkey can be included in order to increase the generalizability of the results. Further, apart from primary school teachers teaching to 6^{th}, 7^{th}, and 8th grades; teachers teaching in elementary schools (grades from first to fifth), in high schools (grades from ninth to twelfth), and in private schools should also be included in order see the main differences among these groups. In addition to the subjects (Turkish, English, Mathematics, Science and Technology, and Social Studies) included in this study, the variety of teaching subjects should also be included in order to compare them with others in this study.

In addition to the teachers, students and other stakeholders' conceptions should also be examined in order to connect the findings to have a more concrete picture of the assessment users' conceptions.

Since an original scale to gather data was not developed by the researchers in this study, original scales or questionnaires having stronger theoretical background should be developed to decrease the number of issues in adaptation process.

The short version of the original scale was conducted to gather data, but the long version of the scale should be adapted and conducted in future studies in Turkey.

The scale to gather data was adapted from English to Turkish, and therefore two items were deleted because of their lower reliability coefficients. The reasons for excluding these items should be discussed in detail in future studies by giving appropriate arguments for these items.

In addition to Exploratory Factor Analysis, Confirmatory Factor Analysis should be conducted to check for the suitability of the model proposed by the researcher of the original scale.

Finally, although there are various teacher conceptions of assessment, this study focused on four main conceptions of assessment. In future studies, some other conceptions related with assessment should be investigated.

Finally, other stakeholders' and assessment users' (e.g. students, parents, teachers trainees..etc.) conceptions of assessment should also be examined in further studies.

REFERENCES

Acar, M., & Anıl, D. (2009). Sınıf öğretmenlerinin performans değerlendirme yöntemlerini kullanabilme yeterlikleri , karşılaştıkları sorunlar ve çözüm önerileri. *TÜBAV Bilim Dergisi, 2* (3), 354- 363.

Anderson, J., White, P., & Sullivan, P. (2005). Using a schematic model to represent influences on, and relationships between teachers' problem-solving beliefs and practices. *Mathematics Education Research Journal, 17*(2), 9-38.

Arda, D. (2009). *İlköğretim sınıf öğretmenlerinin 2005 öğretim programı ekseninde ölçme ve değerlendirme alanındaki yeterlilik ve görüşlerinin incelenmesi*, Unpublished master's thesis, Marmara University, İstanbul, Türkiye.

Bal, A.P. (2009). *İlköğretim beşinci sınıf matematik öğretiminde uygulanan ölçme ve değerledirme yaklaşımlarının öğretmen ve öğrenci görüşleri doğrultusunda değerlendirilmesi.* Unpublished doctoral dissertation. Çukurove Üniversitesi, Adana, Türkiye.

Bennet, R. E., & Gitomer D. H. (2009). Transforming K-12 assessment: Integrating Accountability testing, formative assessment and professional support (pp. 43-61). In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in 21st century*. Mt Gravat, Australia: Springer.

Berry, R. (2008). *Assessment for learning*. Hong Hong: Hong Hong University Press.

Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers et al. (Eds), *Optimizing new modes of assessment: In search of qualities and standards,(* pp. 13-36). Netherlands: Kluwer Academic Publishers.

Black, P., & William, D. (1998). Assessment and classroom learning. *Educational Assessment: Principles, Policy, and Practice, 5*(1), 7-74.

Brookhart, S.M., & Nitko, A.J. (2008). *Assessment and grading in classrooms.* New Jersey: Pearson.

Brown, G.T.L. (2002). *Teacher conceptions of assessment.* Unpublished doctoral dissertation, University of Auckland, New Zealand.

Brown, G. T. L. (2004). Teachers' conceptions of assessment: implications for policy and professional development. *Assessment in Education, 11*(3), 301-318.

Brown, G.T.L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students.* New York: Nova Science Publishers, Inc.

Brown, G.T.L., & Hirschfeld, H.F. (2008). Students conceptions of assessment: Links to outcomes. *Assessment in Education: Principle, Policy, and Practice, 15* (1), 3-17.

Ceylandağ, F. R. (2009). *Teacher self-efficacy beliefs towards measurement and evaluation practices.* Unpublished master's thesis, Middle East Technical University, Ankara, Türkiye.

Cohen, D.K. and Hill, H.C. (2000). Instructional policy and classroom performance: The mathematics reform in California, *Teacher's Collage Record,* 102, 294-343.

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.

Çelikkaya, T., Karakuş, U.,& Öztürk Demirtaş, Ç. (2010). Sosyal bilgiler öğretmenlerinin ölçme- değerledirme araçlarını kullanma düzeyleri ve karşılaştıkları sorunlar. *Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi, 11*(1), 57-76.

Danielson, C. (2008). Assessment for learning: For teachers as well as students. In

C.A.Dwyer (Ed.), *The future of assessment: Shaping teaching and learning*(pp. 191-213). New York: Taylor & Francis.

Degbey, M. (2009). *Using assessment to improve teaching and learning in upper secondary schools*. Unpublished master's thesis, University of Turku, Finland.

Dochy F., Gijbels D., & Segers M. (2006). Learning and the emerging new assessment culture. In L. Verschaffel, F. Dochy, M. Boekaerts, and S. Vosniadou (Eds.). *Instructional Psychology: Past, present, and future trends: sixteen essays in honor of Erik De Corte (pp.191-205).*Netherlands: Elsevier.

Erdal, H. (2007). *2005 ilköğretim matematik program ölçme ve değerlendirme kısmının incelenmesi (Afyonkarahisar ili örneği).* Unpublished master's thesis. Afyon Kocatepe Üniversitesi, Afyon, Türkiye.

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

Field, A. (2005). *Discovering statistics: Using SPSS for windows.* London: Sage Publications.

Fraenkel, J.R., & Wallen, N.E. (2006). *How to design and evaluate research in Education.* New York: McGrow-Hill Inc.

Gay, L. R., & Airasian, P. (2003). *Educational research: Competencies for analysis and applications (7th ed.).* Upper Saddle River, NJ: Pearson Education.

Gelbal, S. & Kelecioğlu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,* 33, 135-145.

Ginsburg, H. P., Jacobs, S. F., & Lopez, L. S. (1993). Assessing mathematical thinking and learning potential in primary grade children. In M. Niss (Ed.)

Investigations into assessment in mathematics education (pp. 157-167). Denmark: Kluwer Academic Publishers.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and teaching: theory and practice, 8*(3), 381-391.

Green, S. B., Salkind,N. J., & Akey, T.M. (2000). *Using SPSS for windows: Analyzing and understanding data.* New Jersey: Prentice Hall.

Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking skills.* Needham Heights, MA: Allyn & Bacon.

Harlen, W. (2007). *Assessment of learning.* London: Sage Publications.

Harlen, W. (2008). Trusting teachers' judgment. In S. Swaffield, *Unlocking assessment* (pp. 138-153). Abingdon, OX: Routledge.

İşler, I. (2009). *Teachers' perceived self-efficacy beliefs and perceptions regarding the implementation of 2004 primary mathematics curriculum.* Unpublished master's thesis, Middle East Technical University, Ankara, Türkiye.

Kaynak, S. (2000). *Ortaöğretimdeki branş öğretmenlerinin öğrenci başarısını ölçme ve değerlendirmeyle ilgili görüşlerinin değerlendirilmesi.* Unpublished master's thesis, Mustafa Kemal Üniversitesi, Hatay, Türkiye.

Kilmen, S. and Çıkrıkçı Demirtaşlı, N. (2009).The perceptions of primary school teachers about their application levels of measurement and evaluation principles. *Ankara University Journal of Faculty of Educational Sciences, 42* (2), 27-54.

Koç, Y., Işıksal, M., & Bulut, S. (2007). Elementary school curriculum reform in Turkey. *International Education Journal, 8*(1), 30-39.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*(4), 4-16.

Marzano, R. J., Brandt, R.S., Hughes, C.S., Jones, B. F., Presseisen B.Z., Rankin, S.C., and Suhor, C. (1988). *Dimensions of thinking: A framework for curriculum and instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.

Milli Eğitim Bakanlığı [MEB](2005). *İlköğretim matematik dersi öğretim programı 1-5. sınıflar.* Ankara: MEB.

Musial, D., Nieminen, G., Thomas, J., & Burke, K. (2009). *Foundations of meaningful educational assessment.* New York: Mc Graw-Hill.

National Research Council [NCR] (2001). *Knowing what students know: The science and design of educational assessment.* Pellegrino, J., Chudowsky, N., and Glaser, R. (Eds.). Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Nunnally, J.O. (1978). *Psychometric Theory.* New York: McGraw-Hill.

Özdemir, S. M. (2009). Sınıf öğretmenlerinin yeni ilköğretim programlarının ölçme ve değerlendirme süreçlerinde karşılaştıkları sorunların incelenmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 42*(2), 55-79.

Sağlam- Arslan, A., Devecioğlu- Kaymakçı, Y., & Arslan, S. (2009). Alternatif ölçme-değerledirme etkinliklerinde karşılaşılan problemler: Fen ve teknoloji öğretmenleri örneği. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 28, 1-12.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research,* 62, 307-332.

Pallant, J. (2007). SPSS survival manual: A step by step guide to data analysis using SPSS for windows (3rd ed.). New York: Open University Press.

Pellegrino, J.W.& Hickey D.T. (2006). Educational assessment: towards beter alignment between theory and practice. In L. Verschaffel, F. Dochy, M. Boekaerts, and S. Vosniadou (Eds.). *Instructional Psychology: Past, present , and future trends: sixteen essays in honor of Erik De Corte (pp.169-189).* Netherlands: Elsevier.

Philippou, G. and Christou, C. (1997). Cypriot and Greek primary teachers' conceptions about mathematical assessment. *Educational Research and Evaluation, 3(*2), 140-159.

Pilten, P. (2001). *Sınıf öğretmenlerinin eğitimde ölçme ve değerlendirme alanındaki anlayış ve uygulamalarının değerlendirilmesi.* Unpublished master's thesis, Selçuk Üniversitesi, Konya, Türkiye.

Sangster, M. & Overall, L. (2006). *Assessment: A practical guide for primary teachers.* Great Britain: MPG Books.

Stiggins, R. (1997). *Student-centered classroom assessment(2nd ed.).*New Jersey: Prentice Hall.

Stiggins, R. J. (2002). Where is our assessment future and how can we get there from here? In R. W. Lissitz & W.D. Schafer (Eds.), *Assessment in educational reform: Both means and ends.* (pp.18- 48) Boston: Allyn& Bacon.

Stiggins, R. and Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory into Practice, 44*(1), 11-18.

Struyf, E., Vandenberghe, R.,& Lens, W. (2001). The evaluation practice of teachers as a learning opportunity for students. *Studies in Educational Evaluation,* 27, 215-238.

Tabachnick, B.S., & Fidell, L.S. (2007). *Using multivariate statistics (5th ed.).* Boston: Pearson.

Thompson, A.G. (1992). Teachers' beliefs and conceptions: A synthesis of the

research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp.127-146). New York: McMillan.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, DC: American Psychological Association.

Uçar, N. (2007). *Elementary school teachers' views about their implementation of the assessment techniques recommended in the new mathematics curriculum.* Unpublished master's thesis, Middle East Technical University, Ankara, Türkiye.

Van de Vijver, F.J.R., & Poortinga Y.H. (2005). Conceptual and Methodological Issues in Adapting Tests. In R.K. Hambleton, P.F. Merenda, C.D. Spielberger (Eds), *Adapting educational and psychological tests for cross-cultural assessment,* (pp.39-63). New Jersey: Lawrence Erlbaum Associates.

Weeden, P., Winter, J., & Broadfoot, P. (2002). *Assessment: What's in it for schools?* Great Britain: RoutledgeFalmer.

Wickline, L. E. (1971). Educational accountability. In E. W. Roberson (Ed.), *Educational accountability through evaluation* (pp.7-18). New Jersey: Educational Technology Publications.

William, D. (2001). An overview of the relationship between assessment and the curriculum. In D. Scott, *Curriculum and assessment* (pp. 165-181). Westport, CT: Ablex Publishing.

Winterbottom, M., Brindley, S., Taber, K. S., Fisher, L. G., Finney, J., Riga, F. (2008). Conceptions of assessment: trainee teachers' practice and values. *The Curriculum Journal, 19* (3), 193-213.

APPENDICES

APPENDIX A

Permisson Letter from MoNE

T.C.
ANKARA VALİLİĞİ
Milli Eğitim Müdürlüğü

BÖLÜM : İstatistik Bölümü
SAYI : B.B.08.4.MEM.4.06.00.06-312/79299 03./05/2010
KONU : Araştırma İzni
Emine VARDAR

ORTA DOĞU TEKNİK ÜNİVERSİTESİNE

İlgi : a) MEB Bağlı Okul ve Kurumlarda Yapılacak Araştırma ve Araştırma Desteğine Yönelik İzin ve Uygulama Yönergesi.
b) Üniversitenizin 06/04//2010 tarih ve 2514 sayılı yazısı.

Üniversiteniz Eğitim Bilimleri Ana Bilim Dalı Yüksek Lisan öğrencisi Emine VARDAR'ın **"Öğretmenlerin eğitiminde ölçme ve değerlendirme hakkındaki kavrayışları. "** konulu tez ile ilgili çalışma yapma isteği Müdürlüğümüzce uygun görülmüş ve araştırmanın yapılacağı İlçe Milli Eğitim Müdürlüğüne bilgi verilmiştir.

Mühürlü anketler (5 sayfa) ekte gönderilmiş olup, uygulama yapılacak sayıda çoğaltılması ve çalışmanın bitiminde iki örneğinin (CD/disket) Müdürlüğümüz İstatistik Bölümüne gönderilmesini rica ederim.

Bülçin UYSAL
Müdür a.
Müdür Yardımcısı

EKLER :
Anket (5 sayfa)

İl Milli Eğitim Müdürlüğü-Beşevler
İstatistik Bölümü
Bilgi İçin: Nermin ÇELENK
Tel : 212 66 40/200—223 75 22
Fax: 223 75 22
istatistik06@meb.gov.tr

APPENDIX B

SAMPLE OF DATA GATHERING SCALE

ÖĞRETMENLER İÇİN "EĞİTİMDE ÖLÇME ve DEĞERLENDİRMEYİ KAVRAMA" ÖLÇEĞİ

Sevgili Öğretmenler,

Değerli görüşlerinize ihtiyaç duyulan bu ölçek; *eğitimde ölçme-değerlendirme sürecini nasıl kavrayıp ne şekilde anlamlandırdığınızı* anlamaya yönelik uygulanmaktadır. **Bu araştırmadaki maddeleri cevaplarken, kendi ÖLÇME VE DEĞERLENDİRME tanımınızı düşünmeniz istenmektedir.**

Bu ölçek; demografik bilgileriniz, ölçme ve değerlendirme dendiğinde aklınıza gelen ölçme ve değerlendirme yöntemlerinin neler olduğu ve ölçme ve değerlendirme süreci ile ilgili görüşlerinizden oluşmaktadır. Araştırmanın amacına ulaşmasında büyük rol oynayacak olan cevaplarınız, sadece araştırma için kullanılacak ve başka bir amaç için kimseyle paylaşılmayacaktır.

Çalışmaya verdiğiniz destek için teşekkür ederim.

Emine VARDAR

ODTÜ, Eğitim Bilimleri Bölümü Yüksek Lisans Öğrencisi

Gazi Üniversitesi, Araştırma Görevlisi

emine_vardar@yahoo.com

BÖLÜM 1: DEMOGRAFİK BİLGİLER

Açıklama: Lütfen, sizin durumunuza en uygun olan seçeneklerin önünde yer alan boşluğa çarpı işareti (X) koyarak belirtiniz.

1.	Cinsiyetiniz: 1. ( ) Bayan 2. ( ) Bay	
2.	Meslekteki Hizmet Yılınız (Lütfen yazınız): .....yıl	
3.	Öğretmenliğe **başlamadan önce**, ölçme ve değerlendirme ile ilgili herhangi bir eğitim aldınız mı? **1.( ) EVET** (Cevabınız "**EVET**" ise **4. Soruya** geçiniz) **2.( ) HAYIR** (Cevabınız "**HAYIR**" ise **5. Soruya** geçiniz)	**4.** Cevabınız "**EVET**" ise, bu eğitimi nerede aldınız? 1. ( ) LİSANS EĞİTİM PROGRAMINDA 2. ( ) LİSANSÜSTÜ EĞİTİM PROGRAMINDA 3. ( ) PEDOGOJİK FORMASYON EĞİTİMİNDE 4. ( ) DİĞER (Lütfen yazınız):.....
5.	Öğretmenliğe **başladıktan sonra**, ölçme ve değerlendirme ile ilgili hizmet-içi eğitim aldınız mı? 1. **( ) EVET** (Cevabınız "**EVET**" ise **6. Soruya** geçiniz) 2. **( ) HAYIR** (Cevabınız "**HAYIR**" ise **7. Soruya** geçiniz)	
7.	Branşınız: 1. ( ) İngilizce 2. ( ) Fen ve Teknoloji 3. ( ) Matematik 4. ( ) Sosyal Bilgiler 5. ( ) Türkçe 6. ( ) DİĞER (Lütfen yazınız):	**8.** Mezun olduğunuz kurum: 1. ( ) Eğitim Fakültesi 2. ( ) Fen-Edebiyat Fakültesi 3. ( ) Diğer (Lütfen yazınız):.....

1 Lütfen, diğer sayfaya geçiniz.

BÖLÜM 2: ÖLÇME ve DEĞERLENDİRME YÖNTEMLERİ

Açıklama:

Aşağıda bazı ölçme ve değerlendirme yöntemleri verilmiştir. Bu yöntemlerden **hangisi ya da hangilerini** ölçme ve değerlendirme sürecinde kullanırsınız?

Size en çok uyan **seçenek veya seçeneklerin** önündeki kutucuklara çarpı işareti koyarak (X) belirtiniz:

❒Uzun cevaplı yazılı yoklama	❒Portfolyo (öğrenci ürün seçki dosyası)
❒Sınırlı cevaplı yazılı yoklama	❒Performans değerlendirme
❒Kısa cevaplı (boşluk doldurma) testleri	❒Akran değerlendirme
❒Çoktan seçmeli test	❒Öz değerlendirme
❒Doğru yanlış testi	❒Grup çalışması
❒Eşleştirme testi	❒Gözlem formu
❒Sözlü sunum	❒Yapılandırılmış grid
❒Sözlü sınav	❒Dereceli Puanlama anahtarı (Rubric)
❒Drama	❒ Diğer (Lütfen yazınız):……..
❒Proje	

BÖLÜM 3: EĞİTİMDE ÖLÇME ve DEĞERLENDİRME SÜRECİ HAKKINDAKİ GÖRÜŞLERİNİZ

Açıklama: Lütfen, ÖLÇME VE DEĞERLENDİRME sürecine ait düşüncelerinize göre, aşağıdaki 25 maddeye ait görüşlerinizi en iyi açıkladığını düşündüğünüz **1 ve 6** arasındaki değerlerden **SADECE BİR TANESİNE** çarpı işareti **(X)** koyarak belirtiniz.

LÜTFEN DİKKAT:

SOL en OLUMSUZ (Hiç Katılmıyorum), SAĞ en OLUMLU (Kesinlikle Katılıyorum) olarak sıralanmaktadır.

HİÇ KATILMIYORUM	(1)	(2)	(3)	(4)	(5)	(6)	KESİNLİKLE KATILIYORUM

		1.HİÇ KATILMIYORUM					6: KESİNLİKLE KATILIYORUM
1.	Ölçme ve değerlendirme sonuçları, okulların sorumluluklarını yerine getirme düzeyleri konusunda ilgili kurum ve kuruşlara bilgi sağlar.	(1)	(2)	(3)	(4)	(5)	(6)
2.	Ölçme ve değerlendirme, öğrencilerin seviyelerine göre gruplara ayrılmasına yardımcı olur.	(1)	(2)	(3)	(4)	(5)	(6)
3.	Ölçme ve değerlendirme, öğrencilerin öğretilenden ne kadarını öğrendiği hakkında bilgi veren bir süreçtir.	(1)	(2)	(3)	(4)	(5)	(6)
4.	Ölçme ve değerlendirme sonuçları, öğrencilere performansları hakkında geribildirim verilmesine yardımcı olur.	(1)	(2)	(3)	(4)	(5)	(6)
5.	Ölçme ve değerlendirme, öğrenme ve öğretme sürecinin vazgeçilmez bir parçasıdır.	(1)	(2)	(3)	(4)	(5)	(6)
6.	Ölçme ve değerlendirme süreci, öğretmenlerin sınıf içinde kullanacakları öğretim yöntem ve tekniklerinde değişiklikler yapmasına imkân sağlar.	(1)	(2)	(3)	(4)	(5)	(6)
7.	Ölçme ve değerlendirme, öğretmenleri inandıklarına aykırı bir biçimde öğretmeye zorlayan bir süreçtir.	(1)	(2)	(3)	(4)	(5)	(6)
8.	Öğretmenler ölçme ve değerlendirme yapsalar bile sonrasında ortaya çıkan bulgulardan çok az yararlanırlar.	(1)	(2)	(3)	(4)	(5)	(6)
9.	Ölçme ve değerlendirme sonuçları, bir okulun kalitesi hakkında bilgi veren geçerli bir göstergedir.	(1)	(2)	(3)	(4)	(5)	(6)
10.	Ölçme ve değerlendirme, öğrencilerin yaptığı çalışmaların not bazında değerlendirilmesini sağlar.	(1)	(2)	(3)	(4)	(5)	(6)
11.	Ölçme ve değerlendirme, öğrencilerin ne öğrendiği hakkında bilgi veren bir süreçtir.	(1)	(2)	(3)	(4)	(5)	(6)
12.	Ölçme ve değerlendirme, öğrenme sürecindeki gereksinimleri hakkında öğrencilere geribildirim verilmesine olanak sağlar.	(1)	(2)	(3)	(4)	(5)	(6)
13.	Ölçme ve değerlendirme sonuçları, öğretim sürecinin şekillenmesine yardımcı olur.	(1)	(2)	(3)	(4)	(5)	(6)
14.	Ölçme ve değerlendirme sonuçları tutarlı olmalıdır.	(1)	(2)	(3)	(4)	(5)	(6)
15.	Ölçme ve değerlendirme, öğrenciler için adil olmayan bir süreçtir.	(1)	(2)	(3)	(4)	(5)	(6)
16.	Ölçme ve değerlendirme sonuçları dosyalanır ancak bunlar ne yazık ki daha sonra tekrar kullanılmayarak dosyalarda öylece kalır.	(1)	(2)	(3)	(4)	(5)	(6)
17.	Ölçme ve değerlendirme sonuçları, okulların yaptığı çalışmaları değerlendirmede kullanılır.	(1)	(2)	(3)	(4)	(5)	(6)
18.	Ölçme ve değerlendirme, öğrencilerin belirlenen hedef ve davranışları kazanıp kazanmadığı hakkında bilgi verir.	(1)	(2)	(3)	(4)	(5)	(6)

19.	Ölçme ve değerlendirme süreci, öğrencinin üst düzey düşünme becerilerinin ölçülmesine olanak sağlar.	(1)	(2)	(3)	(4)	(5)	(6)
20.	Ölçme ve değerlendirme, öğrencilere öğrenme sürecini nasıl verimli geçirmeleri konusunda yardımcı olur.	(1)	(2)	(3)	(4)	(5)	(6)
21.	Ölçme ve değerlendirme süreci, öğrencilerin ihtiyaçları doğrultusunda seçilecek yollarla öğretim yapılmasına olanak sağlar.	(1)	(2)	(3)	(4)	(5)	(6)
22.	Ölçme ve değerlendirme sonuçlarına itimat edilebilmelidir.	(1)	(2)	(3)	(4)	(5)	(6)
23.	Ölçme ve değerlendirme, eğitim ve öğretim sürecinin etkili bir şekilde yürütülmesini engeller.	(1)	(2)	(3)	(4)	(5)	(6)
24.	Ölçme ve değerlendirmenin, öğretim süreci üzerinde etkisi yoktur.	(1)	(2)	(3)	(4)	(5)	(6)
25.	Ölçme ve değerlendirme belirsiz bir süreçtir.	(1)	(2)	(3)	(4)	(5)	(6)

Verdiğiniz destek için teşekkürler.

APPENDIX C

Table Appendix C

Item- total statistics showing the poor items in the first adapted version of the TCoA-IIIA scale

	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	,383	,760
2	,435	,775
3	,408	,779
4	,475	,769
5	,442	,770
6	,441	,740
7	,420	,779
8	,429	,772
9	**,219**	**,833**
10	,348	,767
11	,462	,756
12	,401	,777
13	,422	,768
14	,597	,776
15	,254	,748
16	,335	,780
17	,367	,757
18	**,233**	**,834**
19	,366	,778
20	,357	,779
21	,441	,768
22	,396	,778
23	,468	,755
24	,368	,769
25	,406	,767
26	,408	,777
27	,385	,768

APPENDIX D

Table Appendix D

Total Variance Explained (Initial Factor Extraction)

Component Loadings	Initial Eigenvalues			Extraction SS Loadings			Rotation SS		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6,050	24,201	24,201	6,050	24,201	24,201	4,901	19,606	19,606
2	4,025	16,098	40,299	4,025	16,098	40,299	4,026	16,104	35,710
3	1,581	6,325	46,624	1,581	6,325	46,624	2,082	8,326	44,036
4	1,433	5,734	52,357	1,433	5,734	52,357	2,080	8,321	52,357
5	,992	3,967	56,325						
6	,963	3,852	60,177						
7	,887	3,549	63,726						
8	,819	3,275	67,000						
….	…								
24	,250	1,001	99,470						