DEVELOPMENT OF A TEST FOR ASSESSING TEACHERS' MATHEMATICAL CONTENT KNOWLEDGE FOR TEACHING GEOMETRIC MEASUREMENT AT ELEMENTARY GRADE LEVEL

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

BY

YASEMİN ESEN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF ELEMENTARY EDUCATION

JANUARY 2013

Approval of the Graduate School of Social Sciences

Prof.Dr.Meliha ALTUNIŞIK
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

Prof. Dr.Jale ÇAKIROĞLU
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 Doctor of Philosophy.

Assist. Prof. Dr. Yeşim ÇAPA AYDIN — Co-Supervisor

Assoc. Prof. Dr. Erdinç ÇAKIROĞLU — Supervisor

Examining Committee Members

Prof. Dr. Sinan OLKUN	(ELE, ANKARA U)	
Prof. Dr. Safure BULUT	(SSME, METU)	
Assoc. Prof. Dr. Erdinç ÇAKIROĞLU	(ELE, METU)	
Assist. Prof. Dr. Yeşim ÇAPA AYDIN	(EDS, METU)	
Assist. Prof. Dr. Didem AKYÜZ	(ELE, METU)	

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: Yasemin ESEN

Signature :

ABSTRACT

DEVELOPMENT OF A TEST FOR ASSESSING TEACHERS' MATHEMATICAL CONTENT KNOWLEDGE FOR TEACHING GEOMETRIC MEASUREMENT AT ELEMENTARY GRADE LEVEL

Esen, Yasemin
Ph.D., Department of Elementary Education
Supervisor : Assoc. Prof. Dr. Erdinç ÇAKIROĞLU
Co-Supervisor: Assist. Prof. Dr. Yeşim ÇAPA AYDIN

January 2013, 242 pages

The purpose of this research was to develop and provide evidence for the construct validity of an instrument designed to measure pre-service mathematics teachers' mathematical knowledge for teaching (MKT) measurement specifically on the concepts of length, area and volume.The test is referred as the Test of Mathematical Knowledge for Teaching Measurement (TMK-M).It was aimed to contribute to fill the gaps for lack of valid measures to be used for assessing elementary mathematics pre-service teachers' MKT.

The current test was modeled after the Learning Mathematics for Teaching instruments. Multiple-choice items were constructed to address the portion of the Specialized Content Knowledge and Pedagogical Content Knowledge domains within the pre-determined learning objectives of measurement concepts in Turkish elementary mathematics program.

There were four main rounds of this study: Round One – item development and pilot testing; Rounds Two and Three– field-testing; and Round Four – validation. Participants had been recruited from the departments of elementary mathematics education from almost all districts of Turkey. This participatory study was conducted from the semesters of fall 2010 to spring 2012.

Item and distracter analyses have been conducted to determine item difficulty, discrimination indices and the effect of items on test reliability. Both Classical Test Analyses and Rasch Analyses were conducted in order to see how items functioned and to determine greater number of problematic items.

Keywords: Elementary Mathematics Education, Teacher Knowledge, Pedagogical Content Knowledge, Test Development, Measurement

ÖZ

İLKÖĞRETİM MATEMATİK ÖĞRETMENLERİNİN GEOMETRİK ÖLÇME KAVRAMLARINI ÖĞRETME BİLGİLERİNİ ÖLÇMEYE YÖNELİK TEST GELİŞTİRME

Esen, Yasemin

Doktora, İlköğretim Bölümü

Tez Yöneticisi: Doç. Dr. Erdinç ÇAKIROĞLU

Ortak Tez Yöneticisi: Yrd. Doç. Dr. Yeşim ÇAPA AYDIN

Ocak 2013, 242 sayfa

Bu çalışmanın amacı ilköğretim matematik öğretmen adaylarınınözelikle uzunluk, alan ve hacim ölçme kavramlarını öğretme bilgilerine yönelik çoktan seçmeli bir testgeliştirerek; bu testin geçerlik güvenirlik analizlerini yapmaktır. Bu test literatürde Ölçme Kavramını Öğretme Bilgisi Testi (ÖKÖBT) olarak isimlendirilmiş ve bu alanda literatürde bahsi geçen eksikliklere cevap verebilmek amacı ile hazırlanmıştır.

Test literatürdeki Öğretmek için Matematik Öğrenme Projesi (Learning Mathematics for Teaching) kapsamında geliştirilen öğretmenlik bilgisi modelini temel alarak geliştirilmiştir. Çoktan seçmeli olarak geliştirilen maddelerin matematik öğretim programındaki ölçme kazanımlarına yönelik olarak özel alan bilgisive pedagojik alan bilgisine hitap etmesi hedeflenmiştir. Maddelerin genel olarak hatırlama ve hesaplama becerilerinden çok kavramsal olarak yapılandırılmasına çalışılmıştır. Maddelerin işlerliğini ortaya çıkarmak için taslak sorular birkaç kez uygulanmış, bu aşamalarda farklı veri toplama teknikleri (nitel ve nicel) uygulanmıştır. Yapılan analizler sonucunda istatistiksel sonuçları istendik aralıklarda maddeler sorular yeniden düzenlenmiş veya elenmiştir.

Test geliştirme çalışmaları için 4 etaplı veri toplama süreci planlanmış, birinci etapta soru taslaklarının hazırlanması ve pilot uygulamaları tamamlanmıştır. İkinci ve üçüncü etaplarda soruların alan uygulamaları yapılmış, son aşamada test formunun son hali oluşturulmuştur. Çalışmanın veri toplama süreci 2010 güz döneminden 2012 bahar dönemine kadar sürmüştür. Katılımcılar Türkiye'nin farklı üniversitelerindeki ilköğretim matematik öğretmenliği bölümü 4. sınıf öğretmen adaylarından oluşmaktadır. Maddelerin zorluk ve ayırtedicilik değerlerini belirlemek amacı ile madde ve çeldirici analizleri yapılmıştır. Ayrıca Klasik Test Teorisinin madde analizi araçlarının yanı sıra madde ve kişi bağımsız madde indekslerini analiz etmek amacı ile bir parametreli madde tepki kuramı modeli olan Rasch Analizi yapılmıştır. Bu bağlamda madde uyum indeksleri ve kişilerin yetenek kestirimleri hesaplanmıştır. Ayrıca Rasch Analizi madde zorlukları ve kişi yetenek kestirimlerine dair Klasik Test Teorisine göre daha kapsamlı bir analiz sunmaktadır.

Anahtar Sözcükler: İlköğretim Matematik Eğitimi, Öğretmenlik Bilgisi, Pedagojik Alan Bilgisi, Test Geliştirme, Ölçme.

To everyoneto whomI owe for being “me”

To my other half, Yeşim

To my parents, Cemile and Hasan

ACKNOWLEDGMENTS

This research project came to fruition because of the generosity of many people. Therefore, it is with deep appreciation that I acknowledge those who have assisted and supported me in accomplishing my lofty goal. First I express my thanks to my wonderful family who has helped me through this dissertation process. Without their love and support, I would not have been able to do this. To Yeşim, my wonderful sister, thank you very much for always was being there for me even if it was just a phone call. You are the best sister anyone could ever ask for, even if you are the only one I have, and also thank you for putting up with my late nights and long days and all the complaining and listening to my frustrations, your efforts to calm me down and also owe you to be always aware of how things were going and to provide unconditional support. To my parents: thank you for your love, support throughout all of this. Youhave tenderly fostered my curiosity and love of learning all my life. You always believe in and appreciate me and are always with me when I need. Similar to many experiences in my life, this hard PhD process could not have been completed without your support. To my Grandpa, who was the first person to state trust on my carrier and never gave up doing this from my early ages to college years. I am sure you are seeing and hearing me. To my Grandma, thank you for always working to get me away from school and easing my mind, and "yes this is the time I finished my school (!) ". The rest of my family, uncles and aunts, thank you all for your support and kindness. Additionally, to Zeki Çatal, a person whom I assumed as a member of my family,thank you for all your great help and consistent optimism which enabled me to take every step of this scholastic journey.

I would like to thank my supervisor Dr. Erdinç Çakıroğlu and co-supervisor Dr. Yeşim Çapa-Aydın, for their support and guidance throughout this process. I also would like to thank my dissertation committee members Dr. Sinan Olkun, Dr. Safure Bulut, Dr. Didem Akyüz and Dr. Çiğdem Haser for providing specialized feedback and encouragement.

Along the way, I met and worked with amazing people who have not only acted as colleagues, but as friends, motivators, and helpers. I would like to thank my dear friends; Şule, Jale, Yeliz, Sevgi, Özge Y., Özge E. and Nursel, thank you very much for your support during my PhD journey.

Tuba, thank you for your lovely chats and support and also for calling me to check on me and also my progress.

Kürşat and Dilek, thank you very much for always having a kind word and a smile to share with me.

I also indepted to Oğuz for your support which relieved me and enabled me to continue my progress at a very hard time of my research and also your friendship through this entire process.

I especially want to thank Funda for your encouragement, valuable revisions, contributions, moral support and lovely chats.

Additionally, I would like to thank Murat Abi who helped me for the photocopying works with a smiling face at any time of the day.

I also owe to the endless list of people including the deans, chairs of departments, Faculty members who assisted me in data collection.At this point I wish to express my sincere gratitude to Orhan Ekinci for his kindness and open-ended support during my difficult times.

Last but not the least I appreciate the participants who contributed to this study by spending their time and honestly responding instruments.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

PCK	Pedagogical Content Knowledge
MKT	Mathematical Knowledge for Teaching
SMK	Subject Matter Knowledge
CCK	Common content Knowledge
SCK	Specialized Content Knowledge
KCS	Knowledge of Content and Students
KCT	Knowledge of Content and Teaching
KCC	Knowledge of Content and Curriculum
TMK-M	Test of Mathematical Knowledge for Teaching Measurement
PMT	Preservice Mathematics Teachers
SD	Standard Deviation
M	Mean
N	Sample Size
p	Difficulty Index
D	Discrimination Index
r	Correlation Index

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Literature about teacher assessment shows that teacher assessment tools and examinations have existed since 1850s, and teacher assessment methods have been developed in accordance with changes in the theoretical structures and frameworks. The format of the teacher assessment procedures has been changed, as well as the content. The short and limited personal interview formats that had been utilized in the beginning have evolved into more qualitative and complicated teacher assessment formats in the recent decades. Although there could be different categorizations regarding 80 years of teacher effectiveness research history, Campbell, Kyriakides, Muijs& Robinson (2003) roughly categorize the teacher effectiveness history in four periods according to chronological order. Table 1.1 displays the summary of this information.

Similarly, Hill, Sleep, Lewis, and Ball (2007) summarized teacher assessment history, and reached a similar conclusion about the characteristics and periods of 80 years of teacher assessment history. They pointed out that until the early 1980s, important variables for detecting effective teaching were merely based on either teachers' characteristics, the observable certain behaviors of teachers, or the standardized test scores of either students or teachers. Furthermore, they criticized the shortcoming of teacher assessment perspective -that researchers focused on the observable certain behaviors of teachers, or characteristics of teachers, schools, students, and others to predict the student achievement on standardized tests, which were later found to be too limited and mechanic to cover the complex structure of the classroom environment (Hillet al., 2007).

Table 1.1Periods and characteristics of teacher effectiveness research

Period	Research Characteristics
Presage-product model (1930s–1940s)	the psychological characteristics of teachers were identified and investigated for their effect on learning, including personality (e.g., authoritarianism), attitudes, and experience
Experimental studies (1940s–1960s)	the effects of different teaching styles upon learning were investigated including formal and informal, progressive and traditional, open and closed
Process-product model (1960s–1980s)	the behavior of teachers in classrooms were investigated including the quantity of instruction, focused interactions, and the pacing of instruction, and factors influencing pupil attainment and progress
Teacher knowledge and beliefs model (1990s–present)	teachers' subject knowledge, pedagogical knowledge, and their beliefs such as self-efficacy or expectations were investigated to explore the relationship between these factors and pupil attainment and progress

In 1986, Shulman emphasized that focusing only on isolated behavioral components of teaching masked the main point of teaching, which led to the perception of teaching as a mechanical process. Then he called this aspect as "missing paradigm" in education. He pointed out the lack of teachers' cognitive understanding of subject matter content and that the relationship between such understanding and the instruction teachers provide for students had been overlooked for many decades. Shulman's and his colleagues'contribution was to redirect the focus of teacher effectiveness to the teacher knowledge and the role of content teaching. This approach was quite a radical departure from the conjecture, which focused almost entirely on general aspects of teaching such as classroom management, time allocation, planning, or other general pedagogical issues.The criticisms about deficiency of teacher cognitions in the assessment of teacher effectiveness made researchers to think about more comprehensive conceptualization of teacher knowledge (Carter, 1990; Grossman, 1990; Leinhardt, 1990; Shulman, 1986, 1987). In 1986, Shulman, additionally, shed some light on defining and categorizing teacher knowledge by introducing a new model or new knowledge domain of the teacher knowledge. Shulman (1987) pointed out the existence of the knowledge domain, which differentiates teachers from other adults. Based on Shulman's (1987) characterization of PCK (Pedagogical Content Knowledge), “the category [of teacher knowledge] most likely to distinguish the understanding of the content specialist from that of the pedagogue” (p. 8). Specifically, Shulman (1987) specified PCK as following:

> Special amalgam of content and pedagogy that is uniquely the providence of teachers, their own special form of professional understanding.... It represents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented, and adapted to diverse interests and abilities of learners, and presented for instruction (p. 8).

Scholars working on the teacher knowledge came to an agreement that the knowledge that teacher possesses should be special as any other profession: an engineer, or a physician (e.g. Ball, Lubienski, & Mewborn, 2001; Ball, Hill, & Bass, 2005).According to Loughran, Mulhall, and Berry (2008) “PCK is an academic construct that is based on the view that teaching requires much more than the simple delivery of subject content knowledge to students and, that quality student learning is not the simple recall of facts and figures” (p.93).Behalves of this idea argue that teachers’ mathematical knowledge indeed demands the capability for teaching of mathematics, as differently from capabilities required for the work of mathematicians or other educated adults. Clearly, any adult having basic mathematical background can easily carry out the computation: $\sqrt{3^2+4^2}=\sqrt{5^2}=5$. But in order to handle students’ following wrong response: $\sqrt{3^2+4^2}=\sqrt{3^2}+\sqrt{4^2}=3+4=7$; teaching requires not only recognizing that this student’s answer as wrong but also entails analyzing the case and determining the source of the error. Moreover, error analysis may not be sufficient, so teaching also involves explaining why this is wrong by using different strategies; such as providing counter examples, trial and error or other. At this point, it becomes important to convince the respondent for the next correction step. Finally, for the correction step, teaching involves using multiple representations of the issue. In brief, each step of handling students’ responses involves a deeper and more explicit knowledge of the procedure itself. Each step points to some element of knowledge of concept to teach. Shulman (1986) exemplifies requirements for teachers as following: “the most powerful analogies, illustrations, examples, explanations, and demonstrations – in a word, the ways of representing and formulating the subject that makes it comprehensible for others” (p. 9). Teachers need to know general difficulties of students, for example, in determining the height of the parallelogram, or their tendency of multiplying the given lengths of sides to find the area of parallelogram. At this point, subject matter specialist and any other educated adult may answer any such mathematics question correctly, but teachers are expected to have more than this. For example, they are expected be aware of their students' ideas and their common errors, appropriate teaching strategies etc. Thus, most scholars and policy makers have assumed that such knowledge “not only exists but also contributes to effective teaching and student learning” (Hill, Ball, & Schilling, 2008, p. 372). Moreover, Ball (1991) claims that teacher knowledge have profound effect on all aspects of teaching. Specifically, studies on students' learning and student achievement resulted in a common conclusion that what the teacher knows has a great impact on what is going on the discourse of teaching and what students learn (Fennema&Franke, 1992; Hill, Rowan, & Ball, 2005) .

Although there have been some empirical studies (e.g. Fennema&Franke, 1992; Carpenter, Ansell, Franke, Fennema, &Weisbeck, 1993; Cobb, Wood, &Yackel, 1990) on the relationship between teacher knowledge and student learning, we still have limited knowledge base regarding the determination of professional knowledge and its relationship with student learning. PCK and its components are still ambiguous concepts (Lee &Luft, 2008; Loughranet al., 2008) and need to be studied extensively. Literature suggests that more research is needed to define desired PCK for specific topics and examine its influence on teachers' practices (Kinach, 2002;Park & Oliver, 2007;Segall, 2004;Smith, 1999).

Besides, Ball and her colleagues (2008) emphasize the necessity of improvement in theoretical development of PCK in terms of analytic clarification and empirical testing. They investigate the nature Shulman's (1986) notion of pedagogical content knowledge and propose a practice-based theory of content knowledge for teaching built on PCK (Ball, Thames, & Phelps, 2008). They try to unpack the PCK in their study and propose the following model in order to give in-depth analysis of mathematics knowledge for teaching (Ball et al., 2008). The domain map for “Mathematical Knowledge for Teaching”points out the knowledge the teachers are expected to have. As given in Figure 1.1 the domain map - Mathematical Knowledge for Teaching- has two main subdomains; Subject Matter Knowledge (SMK) and Pedagogical Content Knowledge (PCK).

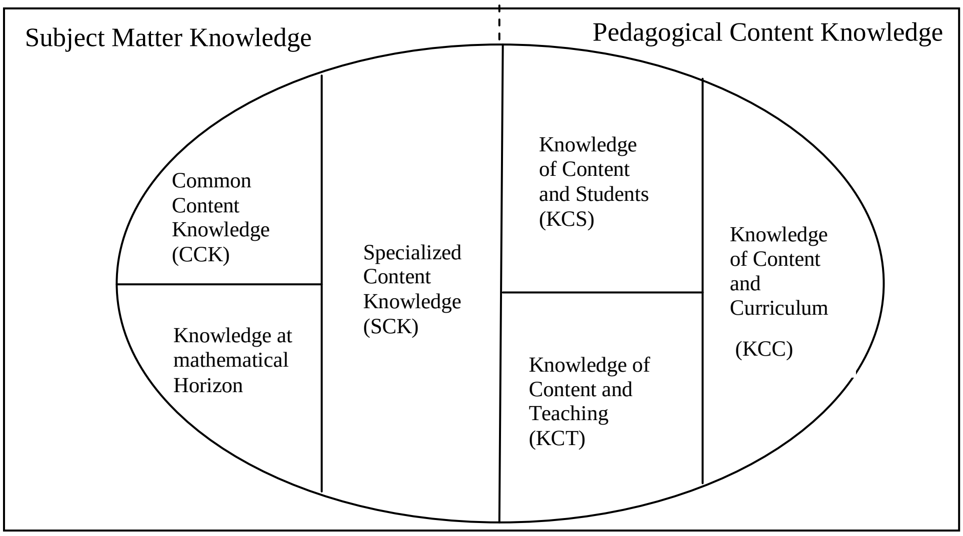


Figure 1.1Domain map ofMathematical Knowledge for Teaching (MKT)

Reprinted from “Content Knowledge for Teaching What Makes It Special” by Ball, Thames and Phelps, 2008, Journal of Teacher Education, 59 (5), p. 389–407

SMK, particularly, covers “how to accurately represent mathematical ideas, provide mathematical explanations for common rules and procedures, and examine and understand unusual solution methods to problems” (Ball, Hill & Bass, 2005, p 378). As seen in Figure 1.1, SMK covers Common Content Knowledge (CCK), which is common mathematical information that also many other professions use. The other knowledge domain that SMK also covers is Knowledge at the Mathematical Horizon. This knowledge allows teachers to see how to link mathematical concepts they are currently teaching to students’ future mathematical learning. Specialized Content Knowledge (SCK), on the other hand, is defined as more than simply a collection of isolated facts and algorithms designed to produce correct answers; instead it also includes a repertoire of interconnected and meaningful concepts and procedures (Ball, 1990). This domain coincides with the Content Knowledge stated in Shulman (1986), which is the part of PCK that refers to the amount and organization of facts and concepts, including an explanatory framework, about a subject in the mind of a teacher, as well as why those facts and concepts are true. Ball (1990) pointed out the differentiation of SCK from the CCK as SCK needed by teachers and non-teachers are actually alike. However, none of those three knowledge domains in SMK – CCK, SCK and Knowledge at Mathematical Horizon contain knowledge related to teaching (Ball et al., 2008).As seen in Figure 1.1, on the right side of the domain map, there exist three knowledge domains related to the Pedagogical Content Knowledge (PCK). These are Knowledge of Content and Students (KCS), Knowledge of Content and Teaching (KCT), and Knowledge of Content and Curriculum (KCC). Compared to the left side of the model, SMK, this part of the model is considered to be more related to teaching profession. In particular, the extent of KCS is summarized as the content knowledge intertwined with knowledge of how students think about, know, or learn this particular content (Hill et al., 2008). Similarly, KCT is the content knowledge related to knowledge of instruction design, the sequence of particular content, instructional advantages and disadvantages of representations used to teach a specific idea and identify what different methods and procedures are necessary during instruction (Ball et al., 2008). Finally, the extent of KCC was also clearly described in the work of Shulman, (1987). He described KCC as the content knowledge related to programs designed for the teaching of particular subjects and topics at a given level considering the program materials in particular circumstances.

Reasons for Assessment of Teacher Knowledge

In the last two decades, identification of teacher qualifications has attracted great attention of policy makers as well as scholars. Hill and her colleagues (2007) summarized the necessity of valid and credible teacher assessment methods in three main themes. The first one is related to political issues. Policy makers have been looking ways to improve and try to find alternative solutions to allocate qualified teachers. But, teachers are one of the key components of curriculum (Fullan, 2001), the qualifications of teachers for implementing the curriculum should be taken into consideration during teacher assessment. Thus, there is a need for valid and credible teacher assessment tools peculiar to the subject matter knowledge for teaching.

The second one is related to academic issues. Scholars have been tried to establish evidences for the effects of teacher education programs on teachers' capacity, knowledge, and skills. For example, teacher education programs have been revised in Turkey as well as the other European Countries in 1998 (YOK, 2007). The content of teaching courses were redesigned to acquire teacher candidates with teaching skills, so it is required to assess teacher candidates through how they are expected to learn. Determining the components of mathematical knowledge for teaching (MKT) –those are also components of PCK- can help us in the process of reforming teacher education programs and can support teachers in developing knowledge required for effective teaching.

The last one is related to the theoretical issues. Several studies have been trying to determine the nature of teacher knowledge. After identifying PCK as a critical component for teaching, researchers and educators tried to find different methodologies and techniques to determine the nature of teacher knowledge, constitutive knowledge domains, and their interrelations between each other, and also they tried to be more precise and accurate in the assessment ofteacherknowledge.Developing such measures are necessary for also validation of the theories. Carrying out an assessment development aiming nationwide will have a positive effect on academic studies. We must recognize that PCK is the heart of the teacher qualification standards that can illuminate the further studies aimed to determine effective teaching. Furthermore, Hill et al., (2008) state that there is lack of information on developing valid and reliable survey measures. There is also need to present development process of content specific measures.

To sum up, based on the reasons stated above the purpose of this study is to develop an instrument for seeking mathematical understandings of teachers, by focusing mostly on teachers' subject matter knowledge for teaching - special forms of mathematical knowledge that are peculiar to the profession of the teaching (Ball, Hill &Bass, 2005; Hill,Bass & Ball, 2005).

There were few projects studying the assessment of teacher knowledge in the literature, which could also provide a clear and well-delineated framework for the purpose of this study. However, considering both the context of Turkey and the purpose of this study, the test was modeled upon the framework of Mathematics Knowledge for Teaching (MKT). For this purpose, four subdomains of Mathematical Knowledge for Teaching; SCK, KCS, KCT, and KCC, which are considered as critically important for teaching profession, took place in the context of this study.

Format of Assessment Tool: Test of Subject Matter Knowledge for TeachingKnowledge

Being aware of complexity of teaching, using single method may be limited for assessing such a complex process. Ideally for a teacher assessment tool to be comprehensive and coherent, all of the components of teacher knowledge should be addressed. The results of authentic and alternative assessment techniques such as, portfolios, case studies, concept maps, group projects, and writing assignments and so on, could provide detailed results and could minimize the standard error of measurement. However, implementation processes require much more time and effort and also many of them are not appropriate for large-scale assessments. Moreover, for a valid assessment, multiple data gathering sources such as observation, interview, and paper and pencil tests should be used. However, when we consider the assessment of mass number of teachers, it becomesalmost impossible to use multiple assessment tools and procedures.

At this point, Haladyna (2004) suggests multiple- choice item format (and its variants) for data collection from large groups based on its effective use and profound research basis. Downing (2006a) confirms the feasibility of multiple-choice item format for large scale cognitive achievement testing. Downing (2006a) emphasizes the effective use of multiple-choice item format for varying range of cognitive taxonomies. Downing (2006b) emphasizes the advantages of selected response item format in terms of efficiency, effectiveness for measurement of cognitive achievement. He attributes the most criticisms on selected response item format to poorly written examples. For this reason, in the context of this study, multiple-choice item format is chosen, in order to be more feasible for data collection from large groups.

Measurement and Geometry

At this point, the instrument in this study covered only specific measurement concepts: length, area, and volume. Through an extensive review of journal articles, professional development series, elementary mathematics curriculum, and surveys of textbooks, the main conclusion was that measurement concepts acknowledged within the mathematics education community as the frequently researched and very important topics in mathematics courses (Fuys, Geddes, & Tischler, 1988; Senk & Thompson, 2003; Simon & Blume, 1994; Thompson & Senk, 2003;) and as topics which students struggle (EARGED, 2003, 2005, 2007; Mullis, Martin, &Foy, 2008; Şişman, Acat, Alpay, & Karadağ, 2011; Van de Walle, 2007). Particularly, there were three main reasons for limiting the test development task to these concepts: (1) length, area, and volume are recurrent concepts for Grades 6-8, (2) these three concepts offer a perspective that helps to get the idea about fundamental structure of mathematical thought, how it evolves: especially deductive reasoning and proof, and (3) students have poor performance of measurement items related to these concepts.

Moreover, measurement is one of the basic tools for students to make sense the world around them. Besides, measurement provides probably the best chance to present students the usefulness of mathematics as well as being an opportunity for motivating students through active learning with realistic problem-solving situations (Lindquist, 1984). Moreover, measurement is one of the main learning domains such as numbers, geometry, algebra, probability and statistics in Turkish Elementary Mathematics Program and measurement provides an opportunity to combine many mathematical concepts within mathematics curriculum such as number, place value, algebra, proportional reasoning, fractions, geometry, data (MoNE, 2008) as well as mathematics with daily life. By the help of measurement skills they can make connection between abstract odor of the mathematics and the way of concrete expression of it (NCTM, 2000).

1.2 Purpose and Problem Statement of the Study

The purpose of this research was to develop and establish the construct validity of an instrument designed to measure pre-service mathematics teachers' mathematical knowledge for teaching measurement. The test is referred as the Test of Mathematical Knowledge for Teaching Measurement (TMK-M). The specific research question is:

1. How valid is the TMK-M?

1.3 Significance of the Study

Most scholars and policy makers have assumed that the Pedagogical Content Knowledge (PCK) not only exists but also has a profound effect on all aspects of teaching (e.g Ball, 1990; Grossman, 1999; Even, 1993; Mason & Spence, 1999; Wilkins, 2008). According to them this is the main knowledge domain for the teaching profession. Even though it is assumed to have a profound effect on teaching, PCK and its components are still ambiguous concepts (Lee &Luft, 2008; Loughran et al., 2008) and need to be studied extensively. In the context of this study, an instrument will be developed that quantitatively measures pre-service elementary mathematics teachers' mathematical knowledge of measurement concepts for teaching, specifically the concepts of length, area and volume. Thus, the nature of PCK on these concepts, its components, and interrelations between components will be investigated.

Understanding of the nature of PCK is important in student learning and students' academic achievement is undeniable. The reviewed literature demonstrates that the PCK and its components need to be carefully examined. It is thought that studying the nature of PCK and its components may enable us to better understand and enhance the student learning in mathematics education. Since there is not much quantitative research examining the nature of PCK in our country, this research may have implications for planning, development, and implementation of teacher education programs aimed to put more emphasis on. By this way, it will be possible to evaluate some of the existing research findings and assumptions regarding pre-service teachers' mathematics knowledge for teaching and how this knowledge can be improved. Therefore, this research will contribute to the body of research that curriculum developers, educators, academicians, and bureaucrats can utilize in developing teacher education programs. This research is important also since understanding and developing mathematics knowledge for teaching may contribute to a shift from a rote learning approach toward a meaningful learning approach. In addition, efforts to improve the content of programs in terms of mathematics knowledge for teaching may also lead pre-service teachers to use this knowledge. It should be noted that all of these possible outcomes may yield a higher level of academic performance in the mathematics education.

As teachers' PCK has a significant effect on teaching performance and all outcomes of the education process, it is necessary to develop assessment tools peculiar to this knowledge. There is a need for presentation of the procedure of developing valid and reliable measures for teachers' knowledge for teaching. The product of this study is one of the examples of these necessary measures. Describing a methodology for developing such an instrument and creating survey items that can be used as a basis in future tools for assessing teachers' PCK were the other purposes of the study in line with the need mentioned.

As is known; the effect of teachers' PCK and its components on teachers' teaching strategies and students' achievement in Turkey has been a new concept and has newly become widely acknowledged by researchers. In other words, no previous study exists in Turkish literature that investigates the teachers' mathematical knowledge for teaching measurement concepts. For this reason, the current study attempted to fill the gap in literature related to the abovementioned topic.

Since teacher education programs have been revised and teacher education courses have been redesigned to acquire teacher candidates with teaching skills, it is required to assess teacher candidates through also how they are expected to learn. Determining the components of mathematics knowledge for teaching (MKT) can help us in the process of reforming teacher education programs and can support teachers in developing the knowledge required for successful teaching.

The other contribution of this study is related to the fact that the determination of teacher qualifications has attracted great attention of policy makers as well as scholars all around the world. Developing measures aimed to assess teachers' MKT is necessary for allocating qualified teachers. Current assessment initiatives are not developed for exclusively assessing teachers' MKT. However, as teachers constitute one of the key components of curriculum (Fullan, 2001), the qualifications of teachers such as MKT should be taken into consideration during teacher assessment. During the teacher assessment, developing a valid and reliable measure is the critical step. Literature suggests that more research is needed to define the desired PCK for specific topics and examine its influence on teachers' practices (Kinach, 2002;Park & Oliver, 2007;Segall, 2004;Smith, 1999). Hence, there is a need for valid and credible teacher assessment tools. There is limited knowledge base for the determination of professional knowledge and its relationship with student learning. At this point it is considered more appropriate to assess teachers' MTK instead of assessing subject matter knowledge or general pedagogical knowledge as done in current application, since, in order to scaffold students during their knowledge construction, teachers are supposed to understand students' conceptions, misconceptions, and learning difficulties. In light of the foregoing, the test development procedure and the product of this study are aiming at establishing a prototype for such kind of measures.

1.4 Basic Assumptions of Study

In the current study, following assumptions were made during creating the measures:

1) Participants' total test scores were resulted from their mathematical knowledge for teaching.
2) Content of the test was based on what teachers usually experience during teaching geometry and measurement in their education.
3) Participants did not receive any other outside help during the administrations.
4) Participants provided their best effort on the items, gave honest and accurate information on test items.

1.5 Definitions of Terms

Content Knowledge: It is referred as subject matter knowledge including the understanding of key facts, concepts, principles and frameworks in a discipline as well as the rules, procedures, proofs, and underlying ideas within that discipline (Brown & Borko, 1992).

Pedagogical Content Knowledge (PCK) : One of the commonly used definition of PCK is that the knowledge domain "goes beyond the knowledge of subject matter per se to the dimension of subject matter knowledge for teaching" (Shulman,1986; p.9). The other detailed and commonly used version of PCK definition is:

> …the most useful forms of representation of ideas, the most powerful analogies, illustrations, examples, explanations and demonstrations,… an understanding of what makes the learning of specific topics easy or difficult: the conceptions and preconceptions that students of different ages and backgrounds bring with them to learning. (Shulman, 1986, p.9)

In this study, it is also referred as "content knowledge for teaching mathematics" (Ball, Hill & Bass, 2002).

Curricular Knowledge: It is defined as the "understanding of curricular alternatives available for instruction" (Shulman,1986; p.10).Content knowledge related to programs designed for the teaching of particular subjects and topics at a given level considering the program materials in particular circumstances (Shulman, 1987) .

Mathematics Knowledge for Teaching (MKT) : It is a multidimensional construct that represents the professional knowledge of mathematics needed by teachers. (Ball & Bass, 2000). Specifically, the mathematical knowledge "used to carry out the work of teaching mathematics" (Hill, Rowan & Ball, 2005, p.373).

Knowledge at the Mathematical Horizon: This knowledge allows teachers to see how to link mathematical concepts they are currently teaching to students' future mathematical learning.

Common Content Knowledge (CCK) : Common content knowledge is the mathematical knowledge and skills used in all professions and settings (Ball, Thames, & Phelps, 2008).

Specialized Content Knowledge (SCK) : It is mathematical knowledge that refers to the amount and organization of facts and concepts, including an explanatory framework, about a subject in the mind of a teacher as well as why those facts and concepts are true. It is defined as more than simply a collection of isolated facts and algorithms designed to produce correct answers; instead it also included a repertoire of interconnected and meaningful concepts and procedures (Ball, 1990).

Knowledge of Content and Students (KCS) : It is summarized as the content knowledge intertwined with knowledge of how students think about, know, or learn this particular content and something particular in how students learn (Hill et al., 2008).

Knowledge of Content and Teaching (KCT) : It is the content knowledge related to knowledge of instruction design, the sequence of particular content, instructional advantages and disadvantages of representations used to teach a specific idea and identify what different methods and procedures are necessary during instruction (Ball et al., 2008) .

Knowledge of Content and Curriculum (KCC) : It is clearly described in the work of Shulman, (1987) . He describes KCC as the content knowledge related to programs designed for the teaching of particular subjects and topics at a given level considering the program materials in particular circumstances.

Elementary Grades: 6^{th} to 8^{th} grades

Pre-service Elementary Mathematics Teacher: a university student enrolled in a department of elementary mathematics education program who has the intention of teaching in a elementary school mathematics.

1.6 Overview of the Research Design

According to Clark and Watson (1995) the main purpose of the test development is “to create a valid measure of an underlying construct” (p.309) and Downing (2006) attributes the validity of tests to their systematic and detail-oriented approach to provide evidence for each test development step, and sufficient evidence to support the proposed inferences from the test scores. For this purpose, Downing (2006) provides a model of systematic test development steps and he also summarizes content and extent of those steps in his study. The content and extent of the each step test development model provides a general framework for not only this study but also method chapter as well. This section will provide an overview of the development of the test. A brief description of how the instrument was developed is presented in the following subsections. The shortened and adapted guideline of test development steps of this study was presented in Table 1.2. As seen in the Table 1.2, the content and extent of the steps and the section where these steps reported and summarized were given.

Table 1.2Test Development steps in the current study

Steps	*The Content and Extent of Steps*	*Chapter where the steps will be reported*
1. Overall Plan	Systematic guidance for test development steps: construct; desired test interpretations; test format (s) ; major sources of validity evidence; clear purpose; psychometric model; timelines; security; quality control	Chapter I
2. Content Definition	Sampling plan for test domain; essential source of content-related validity evidence; delineation of mathematics knowledge for teaching	Chapter II and Chapter III
3. Test Specifications	Operational definitions of mathematics knowledge for teaching; framework for validity evidence related to systematic, defensible sampling of content domain; item characteristics of multiple-choice format	Chapter II and Chapter III
4. Item Development	Development of effective stimuli; formats; validity evidence related to adherence to evidence based principles; item editing	Chapter III

Table 1.2 (cont)

5. Test Design and Assembly	Designing and creating test forms; selecting items for specified test forms; operational sampling by planned blueprint; pretesting considerations	Chapter III
6. Test Production	Publishing activities; printing; security issues; validity issues; issues concerned with quality control	Chapter III
7. Test Administration	Validity issues concerned with standardization; proctoring; security issues; timing issues	Chapter III
8. Scoring Test Responses	Validity issues: quality control; key validation; item analysis	Chapter IV
9. Reporting Test Results	Validity issues; quality control; timely; meaningful;	Chapter IV

Adapted from "Twelve Steps for Effective Test Development" by Downing, 2006, Handbook of Test Development, Downing & Haladyna (Eds), Mahwah, NJ: Lawrence Erlbaum Associates, p.5.

Since this study was designed for specific concepts of elementary mathematics curriculum and dependently had no purpose of making high stakes decisions. Hence, previously mentioned two of test development steps including passing scores and item banking were ignored in the context of this study. The rest of the steps were accomplished at some level of detail.

1.7 Organization of the Study

The following chapters will present some background information on the impetus for development of the test and an evaluation and analysis of the instrument from multiple psychometric perspectives. Chapter 1 presents general information about test development procedures and signifies the importance and significance of the study by summarizing the related theoretical background. The hypothetical model is introduced in this section as well. Chapter 1 also consists of up with giving the definitions of the important terms used in the current study. Chapter 2 presents relevant literature in three major areas that have impacted the test development process. First, a review of the teacher knowledge and its influence of the education system and student achievement are presented. Assessment teacher knowledge was also discussed with a focus on development procedures, outcomes, and common themes. Next, a focused review of measurement education literature is presented. This literature was consulted to develop items for the test that would include known misconceptions and common student errors. Finally, an overview of test theory methods is presented to lay the framework for the analysis to be presented in further chapters. Chapter 3 describes an overview of the research design; the development process for the test from topic selection to revision practices as well as the major characteristics of the participants, data collection and analysis procedures, and validation issues. Chapter 4 provides a more detailed view of the test. An item-by-item analysis is presented in the form of initial versions of the test. For each item, item analysis statistics from the perspective of both Classical Test Theory and Item Response Theory are reported including discrimination indices, difficulty, and item-total correlations. Chapter 4 also presents analyses of the test using methods from Item Response Theory. Chapter 5 includes conclusions drawn from the results of the study and a discussion for future research, and chapter concludes with the implications, limitations, and suggestions for future research.

CHAPTER 2

LITERATURE REVIEW

The purpose of this study was to develop an instrument quantitatively measure pre-service elementary mathematics teachers' (PMT's) mathematical knowledge of measurement concepts for teaching, specifically length, area and volume. This chapter was organized into four main sections. The first part of the chapter provided an overview of teacher knowledge, and knowledge domains required for teaching, especially focusing on content knowledge and pedagogical content knowledge. The second part provided information about assessment of teacher knowledge, such as historical development of teacher assessment techniques, assessing teachers' pedagogical content knowledge and teacher assessment initiatives. The third part provided a review of research studies related to knowledge of and learning measurement, and difficulties and prevalent misconceptions in measurement, specifically studies related to length, area, and volume measurement.

2.1 TeacherKnowledge

Studies on students' learning and student achievement resulted in a common conclusion that what the teacher knows had a great impact on what is going on during class sessions and what students learn (Fennema & Franke, 1992; Hill, Rowan, & Ball, 2005) . At this point, it has been argued that knowledge of teachers have profound effect on all aspects of teaching (Ball, 1991) .

Carter (1990) defined the teacher knowledge as the total knowledge, which underlies his or her actions. However, this knowledge could not be interpreted as all the knowledge a teacher has actually plays a role in his or her actions (Verloop, Driel, & Meijer, 2001) .In fact, there are various cognitive, affective or psychological factors those have an effect on the behaviors of teachers (Lampert, 2001). However, teacher knowledge is mainly related to the cognitive issues domain of teachers. Moreover, some researchers argue that teachers' effects on student achievement are driven by teachers' ability of understanding and using subject matter in their teaching (Leinhardt & Smith, 1985;Shulman, 1986, 1987; Ball, 1991; Ma, 1999). For example, teachers were expected to be able to explain the idea about doing mathematics, and origin and nature of mathematics, organization of facts, concepts and principles, as well as how and why the concepts were interrelated to various groups of students with different characteristics. For each case, teachers should know an explanatory framework considering those characteristics. Briefly, teachers were expected to have a rich conceptual understanding of the particular subject content that they teach (Loughran, Berry, & Mulhall, 2007) . So, most scholars and policy makers have assumed that "such knowledge not only exists but also contributes to effective teaching and student learning" (Hill, et al., 2008, p. 372).

But, having good quality of content knowledge did not guarantee the success at teaching (Ball, 1991; Ma, 1999).The interaction between the amount of content knowledge that teachers possess and the effectiveness or the quality of teaching was inconsistent (Begle, 1979). Monk (1994) and Monk and King (1994) summarized the result such that students at higher levels benefitted from teachers having more content knowledge, but at lower grades there was no effect on student achievement clearly.

The main point for this finding was that "doing mathematics" was different from "teaching mathematics", and scholars working on the teacher knowledge came to an agreement that the knowledge that mathematics, teachers' mathematical knowledge is different from the work of mathematicians or other educated adults (Ball & McDiarmid, 1990;Ball, 1991; Ball, 1993; Ma, 1999;Ball, Lubienski, & Mewborn, 2001; Ball, Hill, & Bass, 2005;Mason & Spence, 1999;Stylianides & Ball,2008). The main conclusion that could be drawn from the studies on teachers' content knowledge was that content knowledge determines the teaching quality, but does not guarantee teaching effectiveness. Especially, considering the fact that teachers need to communicate their mathematical knowledge with children, teachers' content knowledge in practice situations becomes an interesting area to explore.

At this point, Shulman and his colleagues conducted one of the most significant researches on the teachers' knowledge. In fact, the results of this project shifted the teacher knowledge towards the combination of teaching content and pedagogical skills (Shulman, 1986, 1987). Shulman (1986) introduced the notion of new knowledge domain as Pedagogical Content Knowledge (PCK) and points out the existence of the knowledge domain, which differentiates teachers from other adults (Shulman, 1987) .The most widely accepted and commonly addressed definition of pedagogical content knowledge was stated by Shulman (1987).

> special amalgam of content and pedagogy that is uniquely the providence of teachers, their own special form of professional understanding.... It represents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues arc organized, represented, and adapted to diverse interests and abilities of learners, and presented for instruction (p. 8).

In his study, Shulman (1986) elaborated three subdomains of teacher knowledge in detail. The rest of four subdomains are referred to general aspects of education. According to Shulman (1986) , Content Knowledge was defined as "the amount of an organization of knowledge per se in the mind of the teacher." (p.9). On the other hand, Pedagogical Content Knowledge was defined as the knowledge domain "goes beyond the knowledge of subject matter knowledge per se to the dimension of subject matter knowledge for teaching" (Shulman,1986; p.9). The other knowledge domain was Curricular Knowledge, which was defined as the "understanding of curricular alternatives available for instruction" (Shulman,1986; p.10).

After Shulman's work, researchers had continued to work on teacher knowledge domains and to provide comprehensible relationships between these knowledge domains. Although there was a consensus on the impact of teacher knowledge on student learning, there was no clear consensus on which dimension of teacher knowledge was critical for teaching and student learning. Current study focused on two main knowledge domains, which had also been captured the most attention were subject specific content knowledge and pedagogical content knowledge academically.

2.1.1 Content Knowledge

Content Knowledge is one of the main parts of teacher knowledge, have affect the quality of teaching (Grossman, Wilson, & Shulman, 1989) and as well as the student learning (Fennema & Franke, 1992) . Shulman (1986) defined the content knowledge (Subject Matter Knowledge (SMK) also used synonym for content knowledge) initially as 'amount and organization of the knowledge per se in the mind of the teacher' (Shulman, 1986, p. 9).Even (1993) emphasized the importance of powerful content- specific pedagogical preparation for effective teaching, such that "only content specific pedagogical preparation based on meaningful and comprehensive subject matter knowledge would enable teachers to teach the spirit envisioned in the "Professional Standards for Teaching Mathematics" (p.114).

Ball (1990) also highlights the critical aspect of content knowledge such that teachers should understand mathematics deeply in order to be able to represent mathematics in appropriate and multiple ways, to facilitate and handle student understanding of mathematics. Moreover, Ball (1990) characterizes the substantive knowledge in three fundamental principles besides the knowledge about mathematics as following;

Teachers should know knowledge of concepts and procedures correctly such as definition of trapezoid, definition of measurement, how to measure the area of rectangle.

Teachers should understand the underlying principles and meanings such as underlying idea of measurement of area with irregular units.

Teachers should appreciate and understand the connections among mathematical ideas such as the relation between multiplication and area measurement, length measurement and are or volume measurement. Indeed, the materials teachers use –curriculum materials, textbooks and other materials- do not give part for connections among mathematical ideas adequately.

But Loughran et al., (2007) signify that "PCK is an academic construct that is based on the view that teaching requires much more than the simple delivery of subject content knowledge to students and, that quality student learning is not the simple recall of facts and figures." (p.93).For this reason, content knowledge was valuable for classroom interaction (Ball & Bass, 2000) .

2.1.2 Curricular Knowledge

Bruner (1977) expressed the mission of curriculum as:

> A curriculum is more for teachers than it is for pupils. If it cannot change, move, perturb, inform teachers, it will have no effect on those whom they teach. It must be first and foremost a curriculum for teachers. If it has any effect on pupils, it will have it by virtue of having had an effect on teachers (p.xv).

Curricular Knowledge is one of the main parts of teacher knowledge and is defined as content knowledge related to programs designed for the teaching of particular subjects and topics at a given level considering the program materials in particular circumstances (Shulman, 1987) . Teachers were expected to be able to understand the full range of interventions available to particular context for instruction, to be knowledgeable about the order of topics in the same subject matter area as well as curriculum materials apart from her/his own discipline for the specific grade level (Shulman, 1987) .Specifically, Shulman (1987) exemplifies Curricular Knowledge such that "Understanding materials well for that instruction, the alternative texts, software, programs, visual materials, single concept films, laboratory demonstrations or invitations to enquiry?" (p. 10) as well as "the familiarity with the topics and issues that have been and will be taught in the same subject area during the preceding and later years in school, and the materials that embody them" (p. 10) and "curriculum materials under study by his or her students in other subjects they are studying at the same time" (p. 10).

The curriculum knowledge includes two categories of knowledge; the first one is knowledge of particular subject specific goals and objectives, and the other one is knowledge of curriculum materials in particular circumstances. AlthoughShulman (1986) considered curriculum knowledge as separate knowledge domain, later Grossman (1990) conceived curriculum knowledge as a part of PCK considering the characteristics of PCK that the knowledge domain helps to distinguish the subject matter specialist from the pedagogue (Shulman, 1986).

2.1.3 Mathematical Knowledge for Teaching

Based on Shulman's (1987) characterization of PCK, "the category [of teacher knowledge] most likely to distinguish the understanding of the content specialist from that of the pedagogue" (p. 8), scholars working on the teacher knowledge came to an agreement that the knowledge that teacher possesses should be special as any other profession: an engineer, or a physician (Ball et al., 2005) .Shulman (1986) exemplifies requirements of teachers as following: "the most powerful analogies, illustrations, examples, explanations, and demonstrations – in a word, the ways of representing and formulating the subject that makes it comprehensible for others" (p. 9). Besides, Ball and her colleagues emphasize the necessity of improvement in of theoretical development of PCK in terms of analytic clarification, and empirical testing. They investigate the nature Shulman's (1986) notion of pedagogical content knowledge and propose a practice-based theory of content knowledge for teaching built on PCK (Ball, Thames, & Phelps, 2008) . They try to unpack the PCK in their study and propose the following model in order to give in-depth analysis of mathematics knowledge for teaching (Ball, et al., 2008). Based on this study, researchers defines new construct as an important subdomain of "pure" content knowledge unique to the work of teaching, *specialized content knowledge,* with other two important components of PCK (*knowledge of content and students*, and *knowledge of content and teaching*).

As seen in the Figure 1.1 they divide Mathematical Knowledge for Teaching into two parts; *Subject Matter Knowledge* (SMK) and *Pedagogical Content Knowledge* (PCK).

In this model SMK covers *Common Content Knowledge* (CCK), which is common mathematical information how it is used in many other professions that also use mathematics. The other knowledge domain that SMK also covers is *Knowledge at the Mathematical Horizon*. This knowledge allows teachers to see how to link mathematical concepts they are currently teaching to students' future mathematical learning.

Ball(1990) pointed out the definition of *Specialized Content Knowledge* (SCK), as more than simply a collection of isolated facts and algorithms designed to produce correct answers; instead it also included a repertoire of interconnected and meaningful concepts and procedures. SMK also covers "how to accurately represent mathematical ideas, provide mathematical explanations for common rules and procedures, and examine and understand unusual solution methods to problems" (Ball, Hill & Bass, 2005, p 378). This domain is coincide with the definition of *Content Knowledge* stated in Shulman (1986), which is a part of PCK that refers to the amount and organization of facts and concepts, including an explanatory framework, about a subject in the mind of a teacher as well as why those facts and concepts are true. Ball (1990) pointed out the differentiation of SCK, from the *CCK* as special content knowledge needed by teachers and non-teachers alike. But, those all three knowledge domains in SMK – CCK, SCK and Knowledge at Mathematical Horizon - do not contain knowledge related to teaching(Hill et al., 2008).

In the same model PCKalso includes three knowledge domains as *Knowledge of Content and Students* (KCS),*Knowledge of Content and Teaching* (KCT), and *Knowledge of Content and Curriculum* (KCC), and this part of the model is considered as more related to teaching profession than SMK.

KCS is defined as content knowledge intertwined with knowledge of how students think about, know, or learn this particular content and something particular about learners (Hill et al., 2008) .This definition of KCS is coincides with the definition of PCK in Shulman (1987) "an understanding of what makes the learning of specific topics easy or difficult: the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning of those most frequently taught topics and lessons"(p.9). KCS is focused on teachers' understanding of how students learn particular content, how to scaffold students' learning and how to remedy their errors and misconceptions. Being aware of interconnections and differences of these knowledge domains has important implications on development of assessment tools" (Hill, et al., 2008). According to the definition of KCS, for example, teachers need to know general the difficulties of students for example in determining the height of the parallelogram, or their tendency of multiplying the given lengths of sides to find the area of parallelogram. At this point subject matter specialist and any other educated adult may answer this question correctly, but teachers are expected to be aware of their students' ideas and their common errors etc.

KCTis defined as content knowledge related to knowledge of instruction design, the sequence particular content, instructional advantages and disadvantages of representations used to teach a specific idea and identify what different methods and procedures afford instructionally (Ball et al., 2008) .

KCCis defined as content knowledge related to programs designed for the teaching of particular subjects and topics at a given level considering the program materials in particular circumstances (Shulman, 1987) .

2.1.4 Characteristics of Pedagogical Content Knowledge

The analysis of knowledge domains of teachers is still in field of the study the effect of each domain on practice, interaction of each domain with others took the great attention of the researchers, especially PCK. PCK is assumed as a transformation of at least two constituent knowledge domains: general pedagogical knowledge and content knowledge (Ball & Bass, 2000) . PCK differs from content knowledge and general pedagogical knowledge in such a way that content knowledge is the knowledge held by any content expert and general pedagogical knowledge is the knowledge of experienced teachers, such as knowledge of how to organize a classroom and manage students (Gess-Newsome, 1999) . Moreover, PCK is not a single entity that is the common for all teachers of a given subject area (Loughran et al., 2007) . Although there are various conceptualizations of PCK among scholars (Geddis et al., 1993; Gess-Newsome & Lederman, 1999; Grimmett& MacKinnon, 1992; Grossman, 1990) there is a consensus on the two features of PCK: the first one is that PCK is experiential knowledge and skills gained through experience The other one is that PCK is the combination of knowledge, beliefs, and values that teachers develop in the context of teaching situation (Lee & Luft, 2008; Loughran, et al., 2007; Gess-Newsome, 1999). Based on these features of PCK, pre-service teachers or novice teachers usually are expected having limited or minimal PCK; while experienced teachers are assumed as possessing an integrated and developed understanding of teaching (Lee &Luft, 2008). At this point, integrative and transformative models of PCK could be useful to understand how features of PCK evolve with experience of teachers. Gess-Newsome (1999) used mixture versus compound analogy to explain these two models precisely. Considering the mixture analogy, in the integrative model; the knowledge domains of subject matter knowledge, pedagogical knowledge and knowledge of context exist in discrete forms, like the existence of substances in the mixture without any structural change. According to integrative model, any progress in components of PCK enhances the growth of PCK as a whole (Gess-Newsome, 1999). On the other hand, the transformative model is considered as the unification of the components of PCK, that is, a synthetic knowledge of other knowledge domains, similar to the substances in a compound in chemical process (Gess-Newsome, 1999). After the formation of PCK it becomes impossible to separate PCK from other knowledge domains: PCK from subject matter knowledge or PCK from general pedagogical knowledge. Owing to reciprocal relationship between teaching experience and PCK, scholars tend to attribute inexperienced teachers' dependency on discrete knowledge domains to integrative model, while they explain experienced teachers development of PCK through transformative model (Lee &Luft, 2008).

PCK is still an ambiguous concept (Lee & Luft, 2008;Loughran, et al., 2008) and need to be studied extensively. Literature suggests that more research is needed to define desired PCK for specific topics and examine its influence on teachers' practices (Kinach, 2002; Park & Oliver, 2007; Segall, 2004; Smith, 1999). Briefly, researchers studying on teacher knowledge have been focusing on the determination of the nature of teacher knowledge, constitutive knowledge domains, and their interrelations between each other (Hill et al., 2007) .

2.1.5 Pedagogical Content Knowledge of Pre-Service Mathematics Teachers

The transformative and integrative models of PCK introduced by Gess-Newsome (1999) clarify the distinction between PCK structures of novice and experienced teachers. On one end, there is the Integrative model, which assumes PCK as an intersection of three constructs: subject matter, pedagogy and context. On the other hand, for the Transformative model, PCK is assumed as the transformation of subject matter, pedagogical, and contextual knowledge into a unique form, and this unique form is the only source for their instruction. Gess-Newsome, (1999) points out that there is a structural differencebetween cognition of PMT'sand experienced mathematics teachers in term of knowledge domains and teaching expertise.In the Integrative model three distinct knowledge domains exist separately and integrated in teaching context. So, According to Gess-Newsome (1999), in order to be more efficient "each knowledge base should be well structured and easily accessible" (p.13). On the contrary, in Transformative model, there is unique knowledge domain PCK, and " PCK should be be well structured and easily accessible"(p.13).So, in Integrative model teachers are expected to be more fluid in the integration process for each topic, which is taught, whereas in Transformative model, teachers are expected to have PCK for all topics that are taught (Gess-Newsome,1999) .So it was good to summarize the results of studies on PCK of PMT's in terms of knowledge of content,knowledge of content and teaching, knowledge of content and students, and knowledge of content and curriculum.

Before going ahead, it was also good the state the Grossman's(1991) quotation, which also summarized the fundamental expectation from teachers as following:

> "If teachers are to guide students in their journey into unfamiliar territories, they will need to know the terrain well. Both knowledge of the content and knowledge of the best way to teach that content to students, help teachers construct meaningful representations, representations that

reflect both the nature of the subject matter and realities of students' prior knowledge and skills." (p. 203)

2.2 Knowledge of Content

Borko et al. (1992) argue that without a conceptual understanding of mathematical ideas, teaching mathematics from a conceptual perspective is unreasonable. Eisenhart, Borko, Underhill, Brown, and Jones (1993) characterize conceptual understanding as the knowledge of the underlying structure of mathematics and the relation between concepts, as well as the realization of the various relationships between ideas that facilitate meaningful explanations of mathematical procedures (Eisenhart et al., 1993).Ball (1990) emphasizes the statement that being able to "do mathematics" oneself may not sufficient for helping someone else understand and do mathematics, and she indicates that majority of people are able to perform many kind of calculations without understanding the underlying principles or meaning. The results of the studies, which examines the how prospective teachers' mathematical subject knowledge has been conceptualized, revealed that prospective teachers, even in mathematics major, have weaknesses in understanding, particularly in both substantive and syntactic elements of mathematics (Ball, 1990;Tirosh, 2000) . Ball's study (1990) , qualitative in nature, detected that prospective teachers were rule and procedure dependent, and their ability in doing mathematics (knowing the rule and perform calculation) was sufficient for passing their courses but their non-structured, discrete, and procedure based content knowledge might not insufficient for teaching. Similarly, Goulding, Rowland, and Barber (2002) examines how mathematical subject knowledge has been conceptualized and its relationship with classroom teaching. The results of the study revealed that PMT's have weaknesses in understanding, particularly in the syntactic elements of mathematics, and as result of fragilities in insecure subject knowledge and poor planning and teaching.

At this point, Fuller (1997) compares conceptual understanding of novice teachers with experienced counterparts, qualitative research suggested that experienced teachers possess a greater conceptual understanding of certain mathematical topics than their PMT counterparts, and their interaction with both content and students was different in the favor of teachers having more conceptual understanding of mathematics.

2.2.1 Knowledge of Content and Teaching

In 1989,Leinhardt identified and described the important elements needed for teaching be comparing novice and expert teachers. Comparing with expert mathematics teachers, results of the study revelaed that eventhough novice teachers having rich mathematical background could not be transmitted their knowledge to instruction context. Besides there were three main characteristics of lessons of novice teachers. The first one was that their instruction was fragmented in the nature with long transition between lesson segments. The next one was that novice teachers tended to be confused by missent signals frequently. The last one was that they failed to achieveinstructional goals and appeared to be abondoned with ambigious system of goals.Leinhardt (1989) stated that novice teachers were able to be aware of the failures but they lacked of analytics skills to understand where the failures occurred. Borko, Eisenhart, Brown, Underhill, and Agard, (1992) reported that novice teachers struggle with mathematical content they teach and they are very concerned about their limited pedagogical content knowledge and the impact in classroom. In other words, based on results of the study, they explained that novice teachers are more content and procedure oriented. They may fail to explain underlying structure of mathematics, interrelations of concepts, connections, and relationships between ideas for mathematical procedures. The participant having the strongest mathematics background in the study (Borko, et al., 1992) of failed to take appropriate instructional decisions in front of students, and to generate meaningful explanations and examples. Similarly, Mapolelo (1993) noted that strong mathematics background of some novice teachers could not be observed in their teaching experience. Participants of the study mainly depended on procedural and explanation orientated lecture method. They had difficulty in designing meaningful activities that would enhance conceptual understanding and also failed to encourage the students to connect mathematical concepts. Many of the teaching cases revealed that novice teachers experienced difficulties in capturing student conceptions. In fact, the studies on PCK that novice teachers possess and their classroom practices reveal that strong mathematical background does not guarantee the success of student understanding (Borko, et al., 1992; Mapolelo, 1999; Even, 1993) .

To sum up, Ball and Wilson (1990) points out that PCKpossessed by novice teachers was primarily procedural in content and instruction. The results of some studies confirmed that improvement in of one of PCK domains did not enhance PCK holistically.

2.2.2 Knowledge of Content and Students

Carpenter, Fennema, Peterson, & Carey (1988) interpreted the Shulman's (1986) framework for PCK from the perspective of knowledge of the conceptual and procedural knowledge of students bring to the learning environment such as stages of understanding, conceptions, misconceptions, and difficulties of students as well as the required instructional skills those teachers should handle student cases. Instructional skills of teachers include the ability of teachers to connect what to learn for students to what they already possess, and the knowledge of techniques for assessing understanding of students, diagnosing, and eliminating their misconceptions. Research on knowledge about content and knowledge about students thinking pointed out the value of student thinking on becoming an expert (Ball et al., 2001; Carpenter et al., 1988; Fennema & Franke, 1992) . Tirosh (2000) emphasized that prospective teachers' conception about student knowledge was strongly affected by the subject matter knowledge of them as well as naïve beliefs about mathematics and mathematics instruction. Moreover, results of the Tirosh (2000) showed up prospective teachers who were aware of students' tendencies on specific mathematical concepts were able to use this knowledge in their attempts to understand student thinking, and to diagnose the source of student errors. On the other hand, prospective teachers who were unaware of students' tendencies interpreted the source of error as algorithmic mistake and reading comprehension difficulty. Similarly, Vacc and Bright (1999) examined changes in pre-service teachers' abilities to provide mathematical instruction that was based on students' thinking, besides change in their beliefs about teaching and learning mathematics. According to results of this study, pre-service teachers unable to use students' mathematical thinking in their teaching even though they may acknowledge the tenets of student understanding. Livingston and Borko (1990) compared novice and expert teachers in terms of their understanding students. The remarkable result of this study was that novice teachers were more focused on content and task whereas expert teachers were more concerned about student understanding, making connections. In brief, Tirosh (2000) confirmed the claim of Cooney (1994) such that " it makes no more sense to base teacher education programs on the assumption that teachers *tabula rasa*than to assume that students enter their classrooms void of a wide range of conceptions of mathematics" (p.10) in terms of prospective teachers' conceptions about student thinking.

2.2.3 Knowledge of Content and Curriculum

Basturk and Donmez (2011) studied on PMT's level of curricular knowledge about limit and continuity. The result of the study revealed that Turkish PMTs had very limited and rhetoric (based on internet, interaction with others, daily life experiences etc.) knowledge of curriculum. On the other hand, Mitchell and Williams (1993) compared the expert teachers and novice teachers in terms of their technology and curricular material use in classroom. The results of the study showed that expert teachers use curriculum materials more appropriately, more frequently, and more synthetically than novice counterparts. Eisenhart et al. (1993) also confirmed that expert teachers exhibitedgreater tendency towards usingcurriculum materials, learning tools their instruction.

2.3 Assessment of Teacher Knowledge

Literature about teacher assessment shows that teacher assessment tools and examinations has been existed since 1850s, and has been reforming accordance with changes in theoretical structures and frameworks. The format of the teacher assessment procedures has been changed as well as the content. In the beginning, based on short and limited personal interview formats have evolved into more qualitative and complicated teacher assessment formats in recent decades.

Lee and Luft (2008) summarized 1960s teacher assessment perspective that researchers focused on the characteristics of teachers, schools, students and others to predict the student achievement on standardized tests, which were seen too limited and mechanic to cover the complex structure of the classroom environmentlater. Then Leinhardt and Smith (1985) pointed out the deficiency of 1980s teachers assessment tools that researchers focused on the observations of teaching practices - most widely adapted technique for measuring teachers' mathematical knowledge but without aspects of mathematical knowledge for teaching. Besides observations, second method widely used for investigating mathematical content knowledge is exploring teacher knowledge via mathematical interviews and tasks. But this method is not designed to yield generalizable inferences about individual participants' knowledge just to help scholars to understand the nature of and extend of teachers' knowledge. By 1990s some studies began to focus how mathematical knowledge related to students' math achievement. These attempts showed that teacher effectiveness is related to their subject matter than their pedagogical knowledge. These instruments have been evolving in the extent to which they infer the *quality of mathematics in instruction*, as opposed to the *quality of mathematics instruction.* Starting with the idea to test the relationship between mathematical knowledge and student achievement has been evolved to the relationship between mathematics knowledge for teaching and student achievement.

Considering the necessity of assessment issues; researchers and policymakers have been studying on the development of valid teacher assessment tools. Hill et al., (2007) summarized the necessity of valid and credible teacher assessment methods in three themes: the first one is related to political issues. Policy makers have been looking ways to improve and try to find alternative solutions to allocate qualified teachers. The second one is related to academic issues. Scholars have been tried to establish evidences for the effects of teacher education programs on teachers' capacity, knowledge, and skills. The last one is related to the theoretical issues. Several studies have been trying to determine the nature of teacher knowledge. After identifying PCK as a critical component for teaching, researchers and educators tried to find different methodologies and techniques to assess the impact of PCK in order to be more precise and accurate in the teacher assessment, researchers have developed.

2.4 Assessment of Pedagogical Content Knowledge

In general, assessment techniques of pedagogical content knowledge fall into three main categories, such as paper and pencil tests, concept maps, pictorial representations, interviews, and multi-method evaluations (Baxter & Lederman, 1999). First of all, the paper and pencil tests are composed of Likert-type self report scales, multiple choice items and short answer formats. These methods are mainly used to measure teachers' attitudes and beliefs and teachers' subject specific knowledge. In mathematics, "Learning Mathematics for Teaching Project Team" at University of Michigan have been to do an effort on conceptualization and the development of multiple choice items aiming to assess teachers' mathematical knowledge for teaching (Hill, et al., 2004; 2008). But the results of the studies related to this project pointed out the following two critical issues related to multiple choice formats. First one is related to psychometrics of items; that is critical to examine the construct validity of items, whether assuring the construct validity of items means they really measures what they intended to measure at the beginning. Items may overlook the essence of the PCK which is context specific knowledge and practical knowledge in nature. Second, writing multiple-choice items is much more difficult than other assessment formats (Hill et all, 2008). The second category for the assessing PCK is that concept maps and pictorial representations. This method is widely used to reveal cognitive structures of teachers. Teachers asked to make a brain storming about a specific concept and issue. Then they are asked to draw a picture or a chart to illustrate the relationship between those generated words. For the other method in this category is giving teachers prepared cards containing names of different concepts. And then researchers let teachers arrange those cards to represent best relationship among each other (Baxter and Lederman, 1999). At this point Kagan (1990) suggest that concept maps and pictorial representations are generally used in short-term studies. For this reason using these methodologies are more appropriate for short-term changes and methods may fail to explain extensive understanding of PCK. Another criticism for pictorial and concept map assessment category is that researchers are not completely sure whether picture that teachers draw really represent the underlying ideas of teachers. The third and final category for the PCK assessment is multi-method studies. In this category researchers collect data from different sources for data triangulation: such as interviews, observations, cases, written documents, concept maps etc. Although it provides rich data sources for validation of the study, the main criticism for this category is that it requires huge amount of motivation and energy so it is difficult to make a replication study (Baxter & Lederman, 1999).

After covering the main PCK assessment categories, it would be better to retain following general challenges which are suggested by Kagan (1990) during the PCK assessment process. First, PCK cannot be observed directly. The observed behavior of a teacher may be only small portion of his total knowledge. We never have idea about the teacher's posterior knowledge decided not to use through our observations. So, observations provide only a limited view of PCK. Second, despite the attempts to refine the designs of research which aim to reveal teacher cognition precisely, in some cases teachers may fail to express themselves because of either lack of their expression skills or not being aware of their own cognition. The final and obvious critique about PCK methodologies is that most assessment techniques are qualitative in nature. Design, development, data collection, and analysis part of qualitative studies are really time consuming and require high level of energy and motivation.

2.5 Learning and Teaching of the Measurement Concept

The previous section of this chapter reviewed the constructs associated with teacher knowledge without any content-related concern. The current section aims to present an overview of where, how and why the students struggle and feel difficulty related to the length, area and volume concepts. In each subsection, the results of the studies related to the knowledge of PMTs as well as students in length, area, and volume are also taken into account. Since the main idea of this study was focused on the determination of PMTs' pedagogical content knowledge on measurement concepts, the development of measurement concepts in early years of students was mostly excluded from the literature review. The section will start with a discussion of the importance of measurement from the perspective of mathematics teaching and learning, and the fundamental principles of measurement concepts. Then description of the meaning and value of the measurement will continue with a review of misconceptions and difficulties of students on length, area, and volume measurement.

2.6 The Meaning and Value of Measurement

Related research draws attention to two different ways of understanding or definitions of measurement (Simon & Blume, 1994; Zembat, 2007). These are considering the measurement attribute as a quantity and as an evaluation of quantity. The first one is characterized by the definition of measurement by Osborne (1976 as cited in Zembat, 2007, p.205). He defines measurement as a name of an "entity", which presents a quantitative summary of the content of what is measured. For example, when anybody mentions about quantitative attribute of the lake as "450 km^2", he gives a summary of mathematical analysis on the area of lake. The summary contains the information about the quantity of the attribute (450) and the structure of units used to summarize the proportion of the lake (km^2). On the other hand, measurement is characterized by the definition of Bright (1976 as cited in Zembat, 2007, p.205) as a process of a comparison of the object/situation/event being measured and the same attribute of a given unit of measure. From this perspective, measurement is a name of a scaling activity. This means that measurement is the activity of measuring the length of a rope, or measuring the temperature of a room. Due to the frequent use of the term measurement as an activity, the term "measuring activity" is used synonymously. Due to the variety of the "processes" associated with this second meaning of measurement, it takes place within many mathematical concepts. In either case, this second approach guides us to conceptualize the measurement as involving both the constitution of an attribute as a quantity, and the evaluation of that quantity.

2.7 Measurement in Turkish Curriculum

Measurement is one of the main learning domains such as numbers, geometry, algebra, probability and statistics in Turkish Elementary Mathematics Program collectively (MoNE, 2006). According to the framework of Measurement Objectives (MoNE, 2006), the measurement objectives in the middle grades should focus on intuitive understanding of measurement, understanding what an attribute means, and selecting and using a variety of measurement tools, standard units, measurement systems (either metric or self-created), and the meaning and processes of measurement. The objectives go on to state the importance of estimation in measurement, applying the mathematical techniques, and establishing formulas to determine measurements. Similar to Measurement Standards of the NCTM (2000), students are expected to develop measurement skills for "understanding [of] how these formulas relate to the attribute being measured" (p. 46) at middle grades.

More specifically, in early grades (Grades1-3) of Mathematics Curriculum, students are expected to develop intuitive sense with basic quantifying skills of length, area and liquid volume measurement, and necessary measurement skills in their daily life such as those related to money and time. At fourth grade, students are expected to differentiate the perimeter and area, to perform perimeter and area calculation, and to understand the relation between the units of liquid volume measurement and convert them to each other. Based on the previous knowledge, students are expected to improve their performance on perimeter and area calculations and to develop an intuitive understanding of volume measurement at fifth grade.

On the other hand, at sixth grade, students are expected to review length, area and liquid measurement, and as differently from former measurement concepts, they become acquainted with the angle and angle measurement, but they are not expected to perform angle measurement. At seventh grade, they are expected to perform angle measurement, to calculate perimeter and area of two-dimensional shapes, and to calculate surface area and volume of three-dimensional geometric objects. However at seventh grade students are confronted with new two-dimensional concept; circle, accordingly the perimeter and area of circle, surface area and volume of cylinder.

Finally at eight grade students are expected to review and conclude mathematical formulas on volume and surface area of prisms besides the specific relationships on triangles, such as Pythagoras and Thales. To sum up, understanding area and volume measurement of geometric shapes such as prisms, cubes, and cylinder involves a large portion of measurement in sixth-eighth grades mathematics (MoNE, 2006).Measurement experiences at the elementary grades typically focus on empirical measurement such as calculation of perimeter, area, and volume. Other measurement concepts; such as time measurement is a concern in early measurement activities in early grades, whereas liquid volume and weight measurement are the concerns of mostly science instruction.

2.8 Students' Difficulties with Measurement

After the summary of measurement concepts throughout grades1 to 8 mathematics curriculum, Turkish elementary grade students had more difficulty in geometry and measurement items in not only international examinations such as Trends in International Mathematics and Science Study (TIMSS) 1999-R (Educational Research and Development Department [EARGED], 2003), TIMSS 2007 (National Center for Educational Statistics [NCES], 2007), Programme for International Student Assessment [PISA] 2003 (EARGED, 2005), PISA 2006 (EARGED, 2007), but also national examinations (EARGED, 2003, 2007).Results from the 1999-R, 2007 (TIMSS) show that Turkish fourth and eighth-grade students scored lower in the content areas of geometry and measurement than any other mathematics content area (EARGED, 2003, 2007). Moreover, there was a 7-points decrease in the mean scores of the students in geometry domain in TIMSS 2007 as compared to TIMSS 1999 (Şişman, Acat, Alpay, & Karadağ, 2011). The picture is not significantly better than the results of PISA 2003, where the results showed that more than 75% of 15 year-old students were at most the second level of geometry. There were six levels predetermined for the mathematics achievement, and the mean scores of Turkey is relatively less than the mean scores of the Organization for Economic Co-operation and Development [OECD] countries as a whole (EARGED, 2005).

2.8.1.1 Students' Difficulties with Length

According to the Elementary Mathematics Curriculum (MoNE, 2006), linear/length measurement objectives in the middle grades focus on intuitive understanding of what the length measurement means, visual comparison, direct measurement by using non-standard and standard units, conversions, knowing the abbreviations of length units, performing length and perimeter measurement, and conclusion with mathematical formulas on perimeter. Similar to other measurement activities, understanding of linear measurement requires the understanding of principles such as requirement of standard unit size, iteration of units, numbering of a unit at its end, and partial units for measuring continuous length. Reece and Kamii (2001) point out the role of indirect comparisons in teaching and learning length measurement conceptually. According to them, "indirect comparisons require the ability to make two kinds of mental relationships — transitive reasoning and unit iteration" (p. 356). Transitivity refers to the ability of making use of a third relationship in order to reach a conclusion about comparison of two measures without any direct comparison. Based on their study, they stated that majority of children are not able to develop transitivity ability before the age of seven even if it is explained to them. Unit iteration, on the other hand, involves making a part-whole relationship within each whole. Reece and Kamii (2001) argue that when children become skillful in unit iteration and develop the logic of it, their measurement becomes exact. Results of the study indicate that most of the children become able to iterate a unit of length at about nine years age (Kamii & Clark, 1997). At this point, results of studies on students' understanding of length measurement indicate that students face a series of difficulties related to length measurement as well as other domains of measurement. The major and most reported students' difficulties, mistakes, and misconceptions on length measurement are summarized in Table 2.1.

As seen in Table 2.1, from the beginning of length measurement learning students often have difficulty in iterating units by leaving gaps or overlapping units while they are working with nonstandard units (Lehrer, Jacobson, Kemeny & Strom, 1999; Hiebert, 1981). At this level, they are also mistaken by thinking that different sizes of units can be used for length measurement at the same time (Hart, 1998; Rayn

&Williams, 2007). In other words, students might use two different sized paper-clips to measure the length of an object, then they make a mistake of counting total number of paper-clips. This observation refers fragile and not well-developed understanding of students on the use of units and nature of linear measurement (Bragg & Outhred, 2001).

Table 2.1Summary of studies on students' mistakes and misconception (Length Measurement)

Mistakes and Misconceptions	*Studies*
Iterating units by leaving gaps or overlapping units	Hiebert, 1981; Lehrer, Jacobson, Kemeny & Strom, 1999
Different units can be used for length measurement	Hart, 1998, Rayn &Williams, 2007
Counting hash marks or numbers on a ruler/scale instead of intervals	Boulton-Lewis, Wilss & Mutch, 1996; Bragg & Outhred, 2000; Mullis, Dossey, Owen & Phillips, 1991; Stephan & Clements, 2003
Incorrect alignment with a ruler	Bragg & Outhred, 2000; Stephan & Clements, 2003
Starting from 1 rather than 0	Ellis & Siegler, 1995; Lehrer, Jacobson, Thoyre, Kemeny, Strom, Horvath, Gance et al., 1998)
Focusing on end point while measuring with a ruler	Blume, Galindo & Walcott, 2007;Bragg & Outhred, 2000;Lehrer, 2003
Mixing units of length with other units of measurement	Szilagyi, 2007
Converting units	Rayn &Williams, 2007
Confusing the concept of perimeter with area	Kidman & Cooper, 1997; Tan-Sisman & Aksu, 2009, Ryan & Williams, 2007
Length has overlapping property, can be used for area measurement	Lehrer, Jenkins & Osana, 1998;

Moreover, Strutchens, Martin and Kenney (2003) state that there is significant gap in students' understanding of how scales on formal measuring tools work. Students may fail to conceptualize the meaning of numbers and hash marks on rulers (Bragg & Outhred, 2000; Boulton-Lewis, Wilss & Mutch, 1996; Mullis, Dossey, Owen & Phillips, 1991; Stephan & Clements, 2003). Bragg and Outhred, (2000) revealed that only a small portion of 5th grade students could conceptualize the meaning of numerals on a ruler, and make use informal units to construct personal rulers.

For this reason, students fail to understand the role of zero point of ruler, and they tend to read the point 1instead of starting from origin (Ellis & Siegler, 1995; Lehrer, Jacobson, Thoyre, Kemeny, Strom, Horvath, Gance et al., 1998).The National Assessment of Educational Progress (NAEP) items reveal this gap by asking students to measure lengths not aligned to zero or when the scale to be used has no numbers on it (Wilson & Rowland, 1993). Due to lack of understanding the nature and structure of ruler, students have tendency to focus on the end point without considering the alignment of the ruler (Lehrer, 2003; Bragg & Outhred, 2000). In his study, Hiebert (1981) confirmed the results that a significant number of 9 – 13 year students fail to conceptualize the meaning of iterations with a unit, tiling property -without overlaps or gaps- of length measurement when the quantity is being measured. Data from Mathematics Assessment for Learning and Teaching (MaLT) Project database revealed that students have difficulty in converting the metric length units such as centimeters to meters or vice versa (Ryan & Williams, 2007). As a part of MaLT Project, 11 year-old children were asked to decide the recent height at the time of jumping 20 cm above from the point that is 1 meter high. But 23% of students converted 1 meter 20 centimeters to 12 or 1020 meters, which could be interpreted as conversion misconception, which results from 10 or 1000 centimeters in meter misconception (Ryan & Williams, 2007). In addition to this, students have difficulties of mixing the units of length with other units of measurement (Szilagyi, 2007).

Results of the large-scale assessments showed that there are serious weaknesses on students' understanding of length measurement, indeed their knowledge is fragile and not well developed. Items related to length measurement from large-scale assessments can roughly be categorized into three groups:a) nature of a measurement tool, b) quantifying the length of an object, and c) side length and perimeter relationship of shapes. For example, results of released item that asks 4th grade students to find "length of a toothpick placed on the broken-ruler" in NAEP (National Assessment of Educational Progress) in United States in 2003 revealed that only 20% of them were able to reach the correct answer (Blume, Galindo & Walcott, 2007). Similarly, results of item, which asks 3rd, 4th 7th, and 8th grade students to determine “the length of an object pictured above a ruler when the end of the object and ruler were not aligned”, revealed that majority of students found it hard to reach a correct solution in NAEP 1985-86 (Lindquist & Kouba, 1989), NAEP 1996 (Martin & Strutchens, 2000) and NAEP 2003 (National Center for Education Statistics [NCES], 2003). Comparably, results of 1999 TIMMS (Trends in International Mathematics and Science Study) released item asking 8th grade students to find “the length of a curved string placed on a ruler” showed that less than half of 8th graders were able to answer the item correctly (International Study Center, 2000). Mullis, Dossey, Owen and Phillips (1991) attributed these results to difficulty of students in understanding the nature of measurement tools and how to use these tools. From the perspective of Turkish context, the results of TIMSS 1999-R-R (EARGED, 2003) and TIMSS 2007 (Mullis, Martin, &Foy, 2008) were not pleasant either. Specifically, for the item in TIMSS 1999-R-R, which aimed students to interpret the nature and use of unit in length measurement, asked 8th grade students to determine the person who has “most paces to walk to end hallway”. Disappointing results of this item showed that only 976 of more than7000 students gave an answer to this item and only 45.5% of them answered correctly. The International Average Percentage (IAP) of 8th grade students was relatively high compared to Turkish students and the mean of correct response on this item was 67.3%. The content of the other item was about “accuracy of ruler”, and only 17% (IAP: 48.8%) of Turkish students answered this item correctly among 2937 students who gave an answer. The other two length measurement items in TIMSS 1999-R-R were about “the length of a string pulled straight” and “length of a pipe” respectively. Scores of Turkish 8th grade students on these items were relatively lower than the mean scores of international results. About 2900 students answered both of these two items. However, only 29.5 % (IAP: 41.3%) of them answered correctly for the first one, and only 41.8 % (IAP: 72.3%) of students figured out the true result for the second item.

The other category of length measurement items in large-scale assessments is side-perimeter relationship on the shapes. To illustrate, in NAEP 2007 4th grade students were provided a definition that “perimeter was the distance around”, and were given the length of one side. Then they were asked to find the perimeter of a stop sign given that it had eight sides. But only 43% of 4th graders could correctly calculate the perimeter of the stop sign (NCES, 2007). Along similar lines, released item, including the calculation of perimeter of a rectangle with given side lengths, was asked to 3rd and 7th graders in NAEP 2003 (NCES, 2003). The results of this basic item confirmed the weakness of students on conceptual understanding of perimeter. Only 17% of 3rd graders and 46% of 7th graders responded to this item successfully (Lindquist & Kouba, 1989). On the other hand, the portrait of the result did not differ when students were asked to determine the length of rectangular playground whose perimeter and width were given. Results of this item in NAEP 2005 revealed that most of the 8th graders have a difficulty on calculating the length of one side of rectangle, and only 40% of them reached the correct result (NCES, 2005). The trend of results on side length and perimeter relationship for Turkish students does not change either. Two of two items in TIMSS 1999-R-R were about side length and perimeter relationship. The first one requested 8th graders to find the length of a rectangle from ratio of width and perimeter, the other one requested students to find the ratio of width and perimeter in rectangle conversely. 2926 and 990 students gave an answer to the former and the latter one respectively. Scores of Turkish 8th grade students on these two items were relatively lower than the mean of international scores. Only 31.9 % (IAP: 42%) of students reached true result for the first item, whereas 18.3% (IAP: 21.6%) of them reached true result for the second item respectively. These disappointing scores of students indicate the poor understanding of students on length concept, accordingly poor understanding of perimeter. Martin and Strutchens (2000) confirm the incomplete conceptual understanding of perimeter and length, by interpreting the result of the item, which is asking 4th graders to draw a figure with a given perimeter by the help of ruler in NAEP 1996, since only 19% of those who responded could draw a correct figure.

Another issue that should be revised was about the wrong conceptualization of students of nature of the unit. Data from Mathematics Assessment for Learning and Teaching (MaLT) Project database revealed that 11 year-old students might count diagonal of a unit square as 1 when students were asked to select two shapes with the same perimeter on grid paper. Moreover MaLT Project database revealed that majority of students at varying age levels have area and perimeter confusion.

Concepts of area and perimeter provide probably the best opportunity to present students the usefulness of mathematics through active learning with realistic problem-solving situations. Inherently, these concepts have constituted an important component of mathematics curricula. Nevertheless conceptual meaning and practical use of about perimeter and area were shaded by the overemphasis on the procedural skills and calculations.

2.8.1.2 Students' Difficulties with Area

Area refers to a quantitative measure of two-dimensional surface enclosed by a boundary (Piaget, Inhelder & Szeminska, 1981; Douady & Perrin, 1986; Baturo & Nason, 1996). Hence surface area refers to total amount of all surfaces of a 3D object. Like other measurement concepts, area measurement has a strong relationship with daily life (Hiebert, 1981) but also mathematical concepts, including whole numbers and multiplicative structures (Hirstein, Lamb & Osborn, 1978; Skemp, 1986). The main idea for teaching and learning area is conservation of area as well as the conceptual understanding of area and how to measure it (Piaget et al., 1981). Piaget et al. (1981) emphasize invariance property of area as a prerequisite for area measurement and define conservation as the quantitative value of an area remaining constant after reorganization of it.

According to the most recent measurement objectives (MoNE, 2006), area measurement objectives in the middle grades focus on intuitive understanding of what the area measurement means, visual comparison, direct measurement by using non-standard and standard units, conversions, knowing the abbreviations of area units, perform area and surface area measurement for three dimensional objects, and conclude mathematical formulas on area. Similar to other measurement activities, understanding of area measurement requires the understanding of principles such as requirement of standard unit size, iteration of units, numbering of a unit at its end, and partial units for measuring continuous length. The area formula of "length × width"/ "base × height" is introduced in fourth grade and reviewed repeatedly until 8th grade. Carpenter, Corbitt, Kepner, Lindquist and Reys (1981) pointed out the expected results of overemphasis on the formulas such as discrete and limited knowledge of students on area concept, computation problems, difficulties or misconceptions. There are numerous studies paying particular attention to the difference between of length/linear measurement and area measurement (Battista, 1982; Clements & Stephan, 2004; Kidman & Cooper, 1997; Nitabach & Lehrer, 1996; Nunes et al., 1993; Outhred & Mitchelmore, 1996). They stated that the main difference between these two is the directness of measurement. Despite the direct measurement of length, area is indirectly measured, lengths appearing in the formula for calculating it, in other words measurement becomes more abstract when we shift from length measurement to area measurement.

Research on learning and teaching area measurement indicate that students faceproblems related to understanding of calculations besides conceptualizing the area measurement at almost all grade levels (Fuys, Geddes, & Tischler, 1988; Kidman & Cooper, 1997; Nitabach & Lehrer, 1996; Nunes et al., 1993; Outhred & Mitchelmore, 1996). Results of the studies on teaching and learning of area measurement revealed that not only students but also adults (Baturo & Nason, 1996) and teachers (Tierney, Boyd, & Davis, 1986) have poor understanding of area measurement. These studies usually put the blame on the traditional way of teaching, which depends mostly on the formula instead of conceptual understanding, and overuse of it. The major and most reported students' difficulties, mistakes, and misconceptions on area measurement are summarized in Table 2.2.

As seen in Table 2.2, students feel difficulty with understanding of area concept. Most of the students involved failed to understand the area as a space inside a figure in a study (Hirstein et al., 1978; Maher & Beattys, 1986; Carpenter, et. al, 1981), besides transitivity, and conservation of area as demonstrated in other studies (Hiebert, 1984; Piaget, et. al, 1981). Difficulties on transitivity and conservation of area refer lack of students' understanding on the equivalence of a quantity of area whether the area measurement is for two or more shapes in different forms or reformations of same shape. Students continue to struggle with difficulties about the concept of conservation of area throughout their growing ages.

Table 2.2Summary of studies on students' difficulties, mistakes, and misconceptions (Area Measurement)

Mistakes and Misconceptions	*Studies*
Difficulty in understanding of area as a space inside a figure	Hirstein et al., 1978; Maher & Beattys, 1986; Carpenter, Corbitt, Kepner, Lindquist & Reys, 1981
Difficulty in understandingtransitivity, and conservation of area	Hiebert, 1984; Piaget, Inhelder, & Szeminka, 1981; Kordaki, 2003
Confusing area and perimeter	Ma, 1999; Wilson & Rowland, 1993; Kidman & Cooper, 1996; Kenney & Kouba, 1997; Martin & Strutchens, 2000; Hirstein, Lamb, & Osborne, 1978; Kouba et al., 1988,
Area and perimeter are directly related in that one determines the other	Hart, 1998; Kidman & Cooper, 1996;
Applying the formula for finding the area of a rectangle to plane figures other than rectangles.	Kospentaris, Spyrou & Lappas, 2011;Zacharos, 2006
Difficulty in interpreting the results of the procedure	Doudy & Perrin, 1986, Fuys, Geddis, & Tischler, 1988;Kenney & Kouba, 1997; Lehrer, 2003; Outhred & Mitchelmore, 2000;
Confounding linear and square units	Chappell & Thompson, 1999;Carpenter, 1975; Hiebert, 1981; Maher & Beattys, 1986; Reynolds & Wheatley, 1996; Simon & Blume, 1994; Steffe & Hirstein, 1976; Wilson & Rowland, 1993
Difficulty in understandinginverse relationship between the size of the unit and the number of units	Carpenter et al., 1975; Carpenter & Lewis, 1976

These difficulties even remain with pre-service teachers (Maher & Beattys, 1986; Hart, 1989; Tierney, et. al, 1990). Additionally, students' misconceptions or difficulties related to use of area formula can be categorized into three themes. The first one is where students confuse area and perimeter, and use formulas of perimeter and area interchangeably (Hart, 1998; Kidman & Cooper, 1997). The second one is that they tend to conserve areas by conserving their perimeters and vice versa (Hart, 1998; Kidman & Cooper, 1997) and the last one is that students may apply the formula for finding the area of a rectangle to plane figures other than rectangles.

Moreover results of studies indicated that students have difficulty in interpreting the results of the “length x wide” computation (Doudy & Perrin, 1986, Fuys, et. al, 1988; Kenney & Kouba, 1997; Lehrer, 2003). However, having difficulties in conceptualizing the procedure of area measurementis not the case for only students but also for prospective teachers (Baturo & Nason, 1996; Simon, 1995; Tierney, et. al, 1990).

Difficulties in understanding the concept of area measurement also arise with the concept of the unit of area. Based on the related literature two major categories may help to summarize students’ misconceptions or difficulties related to use of unit. The first one is confounding linear and square units (Carpenter, 1975; Hiebert, 1981; Maher & Beattys, 1986; Reynolds & Wheatley, 1996; Steffe & Hirstein, 1976; Chappell & Thompson, 1999; Wilson & Rowland, 1993; Simon & Blume, 1994; Nunes, et. al, 1993). Specifically Nunes et. al (1993) noted the students’ failure in understanding of parallelism between the attribute of unit and attribute of an object that is measured. They explain the reason of this finding in such a way that students’ early experiences with measure deal solely with length, and according to them students often perceive units of length as being universally applicable. The second one is related to wrong counting strategies of units while performing area measurement (Carpenter & Lewis, 1976). In some cases they tend to count only the whole units neglecting the fragmented units in the area, or counting all parts as a whole unit without considering it is fragmented or not (Carpenter et al., 1975). Moreover Carpenter and Lewis (1976) observed that students have difficulty in understanding the inverse relationship between the size of the unit, and the number of units in when the area is conserved. Outhred and Mitchelmore (2000) discussed reasons underlying students’ difficulty in the area formula that students often have difficulty when they are required to translate their experiential knowledge into mathematical abstraction. Based on their claims, similar to length measurement, there is a one-dimensional and additive action in the process of physically covering a rectangle with unit squares. On the other hand, two-dimensional and multiplicative structure of area formula prevents students from linking their intuitive knowledge and mathematical formalization (Mulligan& Mitchelmore, 1997). Furthermore, the results of studies confirmed that the structure of the rectangular array is not intuitively obvious to children (Battista, Clements, Arnoff, Battista, & Borrow, 1998; Mitchelmore, 1983). At this point, Carpenter, Coburn, Reys, and Wilson (1975) suggested that if enough time is provided to students to develop an understanding of the multiplicative structure of rectangular arrays in class, students' tendency to rote learning the area formula may be prevented.

Data from MaLT Project database revealed that children have difficulty in conceptualizing unit of area through varying age levels. For example, 13 year-old children were asked three items related to the area measurement such as, converting area measure, use of area formula, and calculating the surface area. Initially, calculation of the area of sheet of paper with side length: 210 mm, 297 mm was directed to 13 year-old children, and then they were asked to convert the value of area into square meters (with calculators). MaLT Project database revealed that students have difficulty in converting the metric area units such as square millimeters to square meters or vice versa. Only 3% of children reached a correct result for this item. On the other hand, 31% of children made decimal place error. Besides, only 12% of them used the correct linear ratio (1000:1). The second item was related to the understanding and use of the area formula for rectangles. Children were given two different rectangles, one is 12 x5, and the other is 10 x (…). Then, they were asked to find the unknown length measure by using the information given when those two rectangles have same area. For this item, almost half of children (46%) answered correctly. However, 32% of them calculated perimeter instead of area to find the missing dimension. The last item related to area concept was about the surface area of a compound shape made from cuboids. Similar to previous statistics of items related to area in MaLT Project database, only a small number of children (10%) calculated the surface area correctly. More interestingly, one out of every five children (19%) tended to calculate volume of an object rather than surface area. Carpenter et al. (1988) noted another observation about students' formula reliance on area measurement in a large scale testing in United States that although about half of the 7th grade students could get the correct answer when they were asked to find the area of a rectangle when given both dimensions, majority of students failed to calculate area of a square when given the length of one side. It was interesting that even though majority of students knew that the sides of a square are equal, only 13% of these students applied their knowledge of the area formula.

There are numerous evidences that students also have difficulty in relating the pictorial representation of area and mathematical formulization, even if students were skillful in both processes separately (Mitchelmore, 1983; Outhred, 1993; Doig et al., 1995). This is a similar case for pre-service teachers in the study of Simon and Blume (1994). They asked pre-service teachers to visualize rectangular area as measurable by an array of units. The results of the study showed that despite having a sense of the structure of a rectangular array in a rectangle, pre-service teachers did not seem to conceptualize the relation of array structure, the linear dimensions, and the particular area unit.

Statistics of NAEP item, which asked middle grade students the rectangular area as measurable by an array of units indicated that students have problem conceptualization of area basically. Although students have no problem with counting of units along arrays, they failed to perceive that the given number of units refers the area of rectangular shape. From the Second NAEP (Carpenter, Kepner, Corbitt, Lindquist, & Reys, 1980) 70% of the middle grade students gave the correct answer. Similarly, statistics of the same item from the Fourth NAEP (Lindquist & Kouba, 1989) indicated that only 54% of middle grade students gave the correct answer. The results of these studies confirmed that middle-school students did not assume the unit as a unit of measurement of an area.

Although there has been an overemphasis on the use of formula for area measurement, Carpenter et al (1988) stated that only half of the seventh graders tested could correctly calculate the area of a rectangle labeled with both the length and width.

The statistics of administration of the NAEP (2007) reveals another pinpointed fact that, while some progress has been made, 4^{th} and 8^{th} grade students still have difficulty with concepts related to area (NCES, 2007).From the perspective of Turkish context, the results of TIMSS 1999-R-R (EARGED, 2003) and TIMSS 2007 (Mullis, Martin, &Foy, 2008) were not pleasant either. Specifically, for the item in TIMSS 1999-R-R, there were two items related to the area. The content of those two items were parallel and asked 8th grade students to calculate the area of paved walkway around a pool and the area of path around a garden. Disappointing results of these items showed that about 980 of all students answered these items. For the first one only 27.6% (IAP: 42.3%) of them could answer it in a correct way. Similarly, for the second item there were only 22.4% (IAP: 41.5%) of students answering this item correctly.

Almost all studies on teaching and learning area have indicated that the answer of why students struggle so much on measurement concept -especially area- is hidden under the way of how area concepts are taught in class. The reasons underlying students' difficulty in understanding area concepts could be summarized in six themes, such as; a) early use of formulas-starting from the first steps of introducing area concept (Simon & Blume, 1994), b) not enough allocated time for emphasizing the concept of conservation of area without the use of numbers, c) not enough emphasis on conceptualization of this concept in curriculum or implementation of curriculum, d) fragmented way of studying areas without dynamic relation to their perimeters (Kidman & Cooper, 1997), e) lack of efficiency of manipulative activities in assisting learning (Hart,1993), f) use of units which are preferred to are mainly focus on the introduction area formula rather than conceptualizing area measurement (Douady & Perrin, 1986).

To sum up, related literature strongly recommends that conceptual characteristics of the measurement process in the teaching process should be taken into consideration for an effective teaching and learning of area concept (Battista, 1982; Kidman & Cooper, 1997; Nitabach & Lehrer, 1996; Nunes, Light, & Mason, 1993; Outhred & Mitchelmore, 1996).

Results of studies with pre-service teachers show that approaches of pre-service teachers were not different from students' cases. Tierney, Boyd, and Davis, (1990) observed that pre-service teachers also tend to overrely on the area formula "length x width", incorrectly use this formula with irregular quadrilaterals, and have difficulty in the conceptualizing the area formula and interpreting it. Moreover, pre-service teachers are reported to have trouble in the use of units. Specifically, Simon and Blume (1994) noted that some pre-service teachers tend to use linear units instead of square ones for area measurements. In addition; some pre-service teachers also believe that doubling the lengths of the sides of a square doubles its area (Simon & Blume, 1994; Tierney, et. al, 1990).

2.8.1.3 Students' Difficulties with Volume

Compared to other measurement topics, it is observed that concept of volume has been overlooked in the mathematics education research literature. Rather, it has been investigated more in science education research literature. From the perspective of mathematics education a better, understanding of volume measurement –for specifically teaching volume measurement-requires the understanding of principles such as requirement of standard unit size, iteration of units, numbering of a unit at its end similar to other measurement activities. Volume measurement is one of the comprehensive measurement concepts in elementary mathematics curriculum.

French (2004) defined volume as the amount of three-dimensional space in a solid shape that can be quantified in some manner. More precisely, Piaget and Inhelder (1967) defined three types of volume: a) *occupied volume*, b) *interior volume*, and c) *complementary volume*. Although all these volume definitions produce same measurement value quantitatively, each refers a different perspective about concept of volume qualitatively. For example, *occupied volume* refers the amount of space occupied by an object; on the other hand, *interior volume refers* the 'enclosed volume' or the free space enclosed in a closed surface, in other words capacity of a container. Finally, *complementary volume* refers the volume of displaced liquid when the object submerges in the holder full of liquid.

Although being aware of all these identifications of volume is important for conceptualizing the idea, the characterizations of *occupied* and *complementary* volume are mostly handled in science education, and in mathematics education volume measurement is mostly considered as an *interior* volume.

According to the most recent measurement objectives (MoNE, 2006), volume measurement objectives in the middle grades focus on intuitive understanding of what the volume measurement means, direct measurement by using nonstandard and standard units, conversions, perform volume measurement,knowing the abbreviations of volume units, and conclude mathematical formulas on volume of right prisms, and cylinder.

Related literature identified mental processes that children go through in developing a sense for volume concept (Battista & Clements, 1996; Battista, 1999; Battista, 2003; Battista, 2004). In his studies, Battista strongly emphasizes the role of enumeration of arrays in meaningful learning of not only volume concept but also area measurement. According to researchers the main idea of enumeration may prevent many potential misconceptions and errors even in later grades. Lehrer, Jaslow and Curtis (2003) extended this to other similar type solids and signified the importance of structuring and enumeration of units that need to be fit in a given solid.

Despite its importance for not only mathematics learning but also science education, volume is one of the most difficult concepts for students. Similar to area measurement, the way of initial attempts to introduce the volume concept were mostly based on one dimensional and additive action in physically filling a solid object with unit cubes or any other type of units. On the other hand, three-dimensional and multiplicative structure of volume formula prevents students from linking their intuitive knowledge and mathematical formalization (Simon & Blume, 1994). Furthermore, the results of studies confirmed that the structure of the cubic arrays is not intuitively obvious to children (Battista, Clements, Arnoff, Battista, & Borrow, 1998; Battista, 2004; Mitchelmore, 1983). This encourages students to rely on the use of formulas without conceptual understanding (Enochs & Gabel, 1984), and unfortunately also teachers prefer to explain details of volume concept based on formula (Fuys, Geddes, & Tischler, 1988).

Difficulty and complexity of volume concept have drawn more attention of researchers in the measurement conjecture; consequently researchers have identified several problematic issues for students regarding measurement concepts. The major and most reported students’ difficulties, mistakes, and misconceptions on volume measurement are summarized in Table 2.3.

Table 2.3 Summary of studies on students' difficulties, mistakes, and misconceptions (Volume Measurement)

Mistakes and Misconceptions	*Studies*
Lack of understanding of principles underlying formula	Battista & Clements, 1996, 1998; Battista, 2002; Raghavan, Sartoris, & Glaser, 1998
Difficulty in the ides of conservation of volume	Vergnaud, 1983
Difficulty in counting the cube arrays shown in the diagram	Ben-Chaim, Lappan, & Houang 1985; French, 2004
Difficulty coordinating the numeric models and operations with visual models	Battista & Clements, 1998; Clements, Battista, Sarama, Swaminathan, & McMillen, 1997
Confusion of surface area and volume	Hirstein, 1981
Applying the formula for finding the area of a rectangle to plane figures other than rectangular prisms.	Enochs & Gabel, 1984

Similar to area measurement, relying on the formula results in the lack of understanding of principles underlying formula and also conceptual understanding of surface area and volume measurement (Raghavan, Sartoris, & Glaser, 1998). Battista and Clements (1996, 1998) summarized that students who learned the volume formula without any conceptual understandinghavemore misconceptions and greater difficulties in learning volume concept. At this point, although students can perform the intended calculation, majority of them fail to interpret the results theyreach. Battista (2003) summarized that students are not necessarily aware of the number, which refers required number of units when measuring the area and volume. Moreover, noted that rote memorization and rote application of formulas prevent students to conceptualize the surface area and volume concepts and to improve spatial structuring, Battista (1999, 2002) also mentioned that rote learning impedes students' development of coordinating the numeric models and operations with visual models (Battista & Clements, 1998; Clements, Battista, Sarama, Swaminathan, & McMillen, 1997). French (2004) also confirmed that most of the students confuse surface area and volume and encounter problems while visualizing three-dimensional objects and interpreting their two-dimensional representations. As a matter of fact that Battista and Clements (1998) revealed the teaching of volume and surface area is restricted to a couple of formulas that do not make much sense for children in elementary grades. Specifically, results of their study revealed that teaching of volume formula before the age of children's development of coordinate representations is nonsense. Because they asked 5thgrade students to find the surface area and volume of rectangular prisms composed of unit cubes, but none of the students who used a formula was able to connect the formula to a spatial structuring of the cube building. Enoch and Gabel (1984) emphasized that students have a tendency to use volume formula blindly for the solid objects other than rectangular prisms. They explained this issue as an expected result of common sequence of instruction for teaching the metric system, length, area, and volume accordingly. They believed that the misconceptions held by students might be prevented if there is an alternative sequence of concepts in teaching. Moreover, they suggested that teaching volume might be handled apart from length and area measurement (Gabel & Enochs, 1984).

One of the initial researches on volume concept by Ben-Chaim, Lappan, and Houang (1985) observed that fewer than 50%of middle-grade students could count or estimate the number of cubes (volume) of a rectangular prism structure represented isometrically. Specifically, about 25% of fifth graders 40 (-45% of sixth and seventh graders, and 50% of eighth graders could calculate the correct number of cubes of a rectangular prism. In the same study, Ben-Chaim et al. (1985) identified four categories of common errors made by middle school students such as: a) counting the visible cube faces shown in the diagram, b) counting the visible cube faces and doubling, c) counting the number of cubes shown in the diagram, and d) counting the number of cubes shown in the diagram and doubling. The researchers concluded that these errors result from students' difficulty in spatial visualizing, and interpretation of 2D diagram and how and in what extent it represents the 3D object. The summary of the statistics has not differed throughout the age level of students. The results of the second NAEP revealed that fewer than 40% of 17 year olds could count or estimate the number of cubes (volume) of a rectangular prism structure represented isometrically (Hirstein, 1981).Hirstein (1981) interpreted this result as the confusion between volume and surface area. Similarly, French (2004) confirmed that students have a tendency to confuse the volume and surface area because of the rote memorization of formulas.

From the perspective of Turkish context, the results of TIMSS 1999-R and TIMSS 2007 revealed that Turkish students have difficulties related to the volume concept. Specifically, there were two items related to the volume in TIMSS 1999-R (EARGED, 2003). The content of the first one related to the use of volume formula and asked 8th grade students to find the volume of stacks. The second one is related to the liquid measurement, and asked students to find the number of 250 milliliter-bottleswhen filled with total 400 liters. The statistics related to the first item, which asked students to perform the volume formula, was really disappointing. Indeed only 23.8% (IAP: 55.0 %) of them could answer it in a correct way. Although the difference between scores of 8th grade students and IAP scores is relatively low in the second item which is about liquid volume measurement, the statistics of this item is still worrisome. There was only a 37.9% fraction of students who could reach a correct answer whereas the IAP of this item was 41%.

On the other hand, related research on teachers' understandings of content knowledge related to surface area and volume has shown similar results to students' (Hershkowitz & Vinner, 1984). Mayberry's (1983) study of teacher knowledge with respect to the van Hiele levels revealed that teachers' own understandings of the content resided largely at the first two of the five levels.

To the best of knowledge obtained while reviewing literature, there are a few researches centered on teachers' conceptions about the concept of volume compared to other measurement concepts, such as area, perimeter. Almost all of these studies on volume concepts focus on children. For example, Saiz (2003) investigated primary teachers' conceptions about the concept of volume. Specifically, she asked primary teachers to determine the volume of some daily things, such as a chair, a handkerchief, a sheet of paper, female screw, a spinning top. The results of the study were worrying. The first striking result is teachers' reliance on the formula. This reliance lies behind the many misconceptions of teachers on volume concept. Saiz (2003) revealed that teachers did not consider that objects have volume unless they have apparent third dimension, like sheet of paper or handkerchief. Moreover, teachers did not consider that some daily things have measurable volume due to their irregular shape, like chair, or spinning top. Based on the findings of Saiz (2003), teachers claimed that since the volume formula could not be applied to irregular shapes, the volume of irregular shapes might not exist. To sum up, "in non-scholar situations teachers manifest the usage of different meanings related to the word *volume*; the dominant one is that of volume as a number" (p.100) and "some teachers have constituted a mental object associated with the bodies' characteristic of 'having three dimensions' " (Saiz, 2003, p.100)

Concluding, not only teachers but also researchers should be aware of the difficulties that students face in conceptual understanding of volume, and accordingly in interpreting the procedural applications. Hence, they need to find alternative ways to help students to overcome this problem, but, most importantly, research conjecture should be aware of the difficulties of teachers and should leave the field clear for teachers. Otherwise, it would be too difficult to get out a vicious circle.

CHAPTER 3

METHOD

In this chapter, an overview of the research design, major characteristics of the population and sample, instruments of the study, procedures of data collection and data analysis, and validation issues will be explained.

The first section gives brief information about overview of the research design. The second section clarifies test development design in the current study, which consists of development of test items and field-testing step by step. The third section presents details of the study; the population and data analysis methods. Reliability and validity issues are described in the final section.

3.1 Research Design

The main aim of the current study was to develop an instrument that quantitatively measures pre-service elementary mathematics teachers' mathematical knowledge of measurement concepts for teaching, specifically on the concepts of length, area, and volume. There were four main rounds of this study: Round One – item development and pilot testing; Rounds Two and Three– field-testing; and Round Four – validation. Accordingly, the data were collected at four different time intervals through different methods (i.e. qualitative and quantitative).Participants had been recruited from the departments of elementary mathematics education from almost all districts of Turkey. This participatory study was conducted from the semesters of fall 2010 to spring 2012. Based on this design (Downing, 2006), in this chapter, results of test development procedures and quantitative data analysis of survey data will be explained.

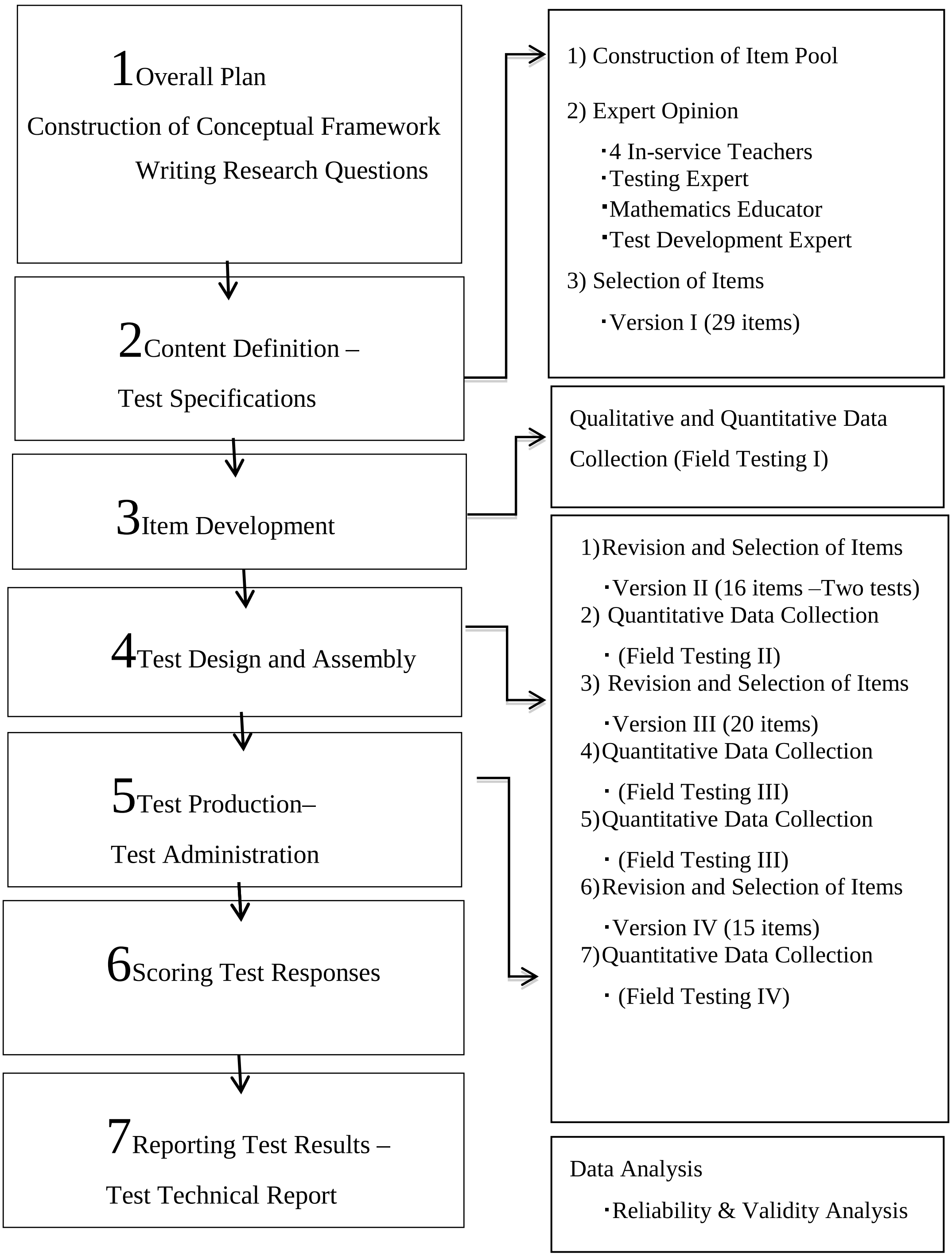


Figure 3.1 Steps followed in the study

3.2 Content Definition of the Test

This study was interested primarily in developing a non-computational multiple-choice instrument that would assess pre-conceptions of pre-service mathematics teachers' mathematical knowledge of measurement concepts for teaching, specifically on the concepts of length, area, and volume. "Mathematical knowledge for teaching" refers to the knowledge domain which teachers are expected to have and which makes difference between mathematics teachers and any other educated adults. Briefly, teachers were expected to have a rich conceptual understanding of the particular subject content that they teach (Loughran, et al., 2007) . To sum up, the purpose of this study was to develop an instrument for seeking mathematical understandings of teachers focus mostly on teachers' subject matter knowledge- special forms of mathematical knowledge that are particular to the profession of the teaching (Ball et al., 2005; Hill et al., 2005). For this reason, four subdomains of Mathematical Knowledge for Teaching; SCK, KCS, KCT, and KCC, which are considered as critically important for teaching profession, took place in the context of this study.

3.3 Preparation of Test Specifications

The second step was to preparethe test specifications and to determine the necessary number of items related to objectives.Determining the test specifications is a similar process in both Classical Test Theory and Item Response Theory. The main point of this step is to set the criteria for conceptualization of the domain (Walsh & Betz, 1995). It is important to define the construct and elaborate the theoretical and empirical basis for the construct; in other words, it is necessary to draw the boundaries and to state its relation with other similar constructs. Drawing boundaries and stating the relations provide relatively specific information regarding the nature of knowledge and how it might be assessed. This step also covers the determination of how measurement should proceed.

Specifically for this study, if there had been well-defined teacher qualifications those were sufficient enough to provide a baseline for teacher assessment tools, there would not be a need to construct content specific table of specifications. When this study started, there were tentative teacher qualifications prepared by MoNE (2008). But these tentative qualifications were too general, thus too weak to provide a baseline for item construction at that time. For this reason, before determining the test specifications, expected learning outcomes in measurement concepts were specified. These learning outcomes were selected because teachers are expected to provide learning environments for students to construct their knowledge on specific measurement concepts. Thus, table of specification was constructed based on the expectations from teachers considering learning objectives in Elementary Mathematics Curriculum in Turkey (MoNE, 2010).

The starting point for selecting topics/themes to cover was mathematics curriculum review for grades 6-8 using content analysis. Table 3.1 summarizes the objectives in Turkish mathematics curriculum review for grades 6-8 (See Appendix E).Curriculum in Turkish elementary mathematics mainly focuses on intuitive understanding of the meaning of measurement, visual comparison, direct measurement by using non-standard and standard units, conversions, knowing the abbreviations of measurement units, performing measurement skills, and it ends with mathematical formulas on measurement (MoNE, 2006). Similar to Measurement Standards of the NCTM (2000), students are expected to develop measurement skills for "understanding [of] how these formulas relate to the attribute being measured" (p. 46) at elementary grades.A list of commonly addressed measurement concepts for all grade levels had been collected and also summarized in Table 3.1. The list was reduced after eliminating uncommon concepts for all grade levels (See Appendix E). Then, four key concepts were identified after examining mathematics curriculum prepared by Ministry of National Education namely; conceptual meaning of measurement in general, length measurement, area measurement, and volume measurement. Table 3.1 summarized these four key concepts.

Table 3.1Summary of key measurement concepts addressed for grades 6-8

Learning Area	*Objectives*
Measurement	Students should be: ▪ explain measurement conceptually. ▪ makes conclusions based on strategically true estimations.
Length Measurement	Students should be: ▪ explain length measurement conceptually. ▪ carry out length measurement computations (such as, length of a line segment, perimeter of circle and polygons) according to their grade level
Area Measurement	Students should be: ▪ explain area measurement conceptually. ▪ carry out area measurement computations (such as, area of polygons, surface area of 3D objects; prisms, cylinder pyramid, sphere and cone) according to their grade level. ▪ explain Pythagoras (Pisagor) theorem conceptually by using area measurement.
Volume Measurement	Students should be: ▪ explain volume measurement conceptually. ▪ carry out volume measurement computations (such as, volume of 3D objects; prisms, cylinder pyramid, sphere and cone) according to their grade level.

Two sub domains of MTK (Specialized Content Knowledge and Pedagogical Content Knowledge) proposed by Hill and her colleagues (2008) were essential components of the test, and each item in the test was classified into one of these two categories. Specialized content knowledge was specified as an essential component of this study since it was one of the main parts of teacher knowledge, has an affect the quality of teaching (Grossman, Wilson, & Shulman, 1989) and as well as the student learning (Fennema & Franke, 1992) .Ball (1990) also highlights the critical aspect of content knowledge such that teachers should understand mathematics deeply in order to be able to represent mathematics in appropriate and multiple ways, to facilitate and handle student understanding of mathematics. For this reason, the weight of specialized content knowledge was equal to pedagogical content knowledge.

3.4 Item Development and Preparation of the Item Pool

The third step for systematic test development is to develop an item pool. This subsection will provide an overview of methods used to systematically develop multiple- choice items. Multiple- choice item format was selected as alternative assessment techniques (portfolios, case studies, concept maps, group projects, and writing assignments) requires much more time and is not feasible for large scale assessments (Haladyna, 2004).

Once the table of specification was constructed, nearly 50 item stems were drafted considering the initial framework of item pool by the researcher. The items were constructed to target subdomains of MKT. At this phase of the test development, researcher conducted an interview with 4 in-service mathematics teachers about the reality of cases for classroom environments. In-service teachers investigate the tool in terms of feasibility of items in class environment, reality and extend of cases given in item stems. The items were intended to be conceptual in nature and each item has one correct answer. At the end of the revision of stems in terms of content and context, about 30 stems were seen as more appropriate to complete, and to construct distracters. Initial distracters were written based on the known difficulties or misconceptions in the literature by researcher. The items had been modified so that each item had total five alternatives just one correct answer. The test is made up entirely of multiple-choice questions that assess understanding of fundamental measurement concepts and those require little or no computation.

Face validity and content validity were determined by revision of two university professors, one test developer, and seven pre-service mathematics teachers. The interest of one professor is mathematics education and the others' is testing. They checked initial format of the items from the perspective of mathematics education and testing principles. Test developer reviewed items from the perspective of test development criteria, especially linguistic features of that may affect comprehension. Before pilot implementation of the test, 7 pre-service mathematics teachers took the early version of the test. They were asked to check the readability, clearness, fluency of items as well as the comprehension of content as being test takers. Some distracters were modified through pre-service teachers' comments on implementation of test. Refer to the Table 3.3 to see an example of initial layout of each item with respect to the categories described in the table specifications. The profile of 29-item table of specifications was displayed in Table 3.3.

Table 3.2Summary of item classification of test

	Sub-Domains of MKT	
	SCK	PCK
The concept of Measurement	MS8, MS20	
Length Measurement	LS2, LS6, LS11, LS12, LS16, LAS25	LAP7, LP28, LAP14, LP29
Area Measurement	AS13, AS19, AS23, LAS25	AP3, AP4,AP5, LAP7, AP10, LAP14, AP15, AP17,AP21, AP22, AVP26
Volume Measurement	VS9, VS18, VS24, VS27	VP1, AVP26

As seen in the Table, there were two items (MS8 and MS20) including specialized content knowledge on the measurement concept in general. These items covered the underlying two ideas on measurement, the first item was about comparison of quantities and the latter one was based on the estimation idea. On the concept of length measurement, there were 10 items (LS2, LS6, LAP7, LS11, LS12, LAP14, S16, LAS25, LP28, and LP29). As summarized in Table 3.3, there were six items on the specialized content knowledge of length measurement, and the rest of them were about the pedagogical content knowledge.

On the other hand, there were totally 15 items on area measurement (AP3, AP4, AP5, LAP7, AP10, AP13, LAP14, AP15, AP17, AS19, AP21, AP22, AS23, LAS25, and AVP26). The frequency distribution of the items on the area measurement was as the following: three of them were on the specialized content knowledge of area concept, three of them included knowledge of content and students of area measurement and the rest of them were about the pedagogical content knowledge of area measurement.

Similarly, there were totally six items (VP1, VS9, VS18, VS24, AVP26, and VS27) on the volume measurement. Four of these items were about the specialized content knowledge on volume concept, and two of them were about the pedagogical content knowledge of teaching volume.

3.5 Test Design and Assembly

Since test development is an iterative procedure, there were four implementation phases and methodology of each implementation will be explained in four sections.

In this study the development and administration of the instrument was interconnected. The researcher decided to use four rounds to administer test items for developing the appropriate multiple-choice items.

- Round One – item development and pilot testing;
- Rounds Two and Three– field-testing;
- Round Four – validation.

Briefly, the data collection of the study began with qualitative and quantitative research methods, which framed the open- ended responses and interview findings. In the first section, both quantitative data collection and interviews were conducted. Other three rounds continued with quantitative data collection (survey research).

Although all 4thyear participants in Turkey were identified as the target population of this study, it is not possible to come into contact with this target population. Thus, accessible population and sampling procedures were determined according to the purpose of each round. Participants had been recruited from the departments of elementary mathematics education from almost all districts of Turkey. This participatory study was conducted from the semesters of Fall 2010 to Spring 2012.

3.6 Context of the Study

Before explaining the sampling and data collection it may be better to give more detail on elementary mathematics teacher education system in Turkey for detailed description of current context of the study. For this purpose information given in report of YOK (2007) (Council of Higher Education) was summarized. The report published in 2007 offers comprehensive information on not only the structure of current teacher education system, but also historical development of teacher education systems from 1982 to 2007.

Although different teacher education models were experienced in the history of Turkey, today all teacher education programs are required to have an undergraduate degree. After the secondary school, there is centralized university entrance examination in Turkey. Like all other undergraduate programs, elementary mathematics teacher candidates are also enrolled in an undergraduate teacher education programs based on their university entrance exam scores. At this point teacher education programs provide two options for teacher candidates. Based on their university entrance exam scores, they can prefer regular class (require higher scores) or night class program. Although coursework of both options is same, there exists a difference only in their course hours. Pre-service mathematics teachers following the former option take their courses at daytime; on the contrary, the course work of the latter option is in the evening. In fact, all teacher education programs in Turkey follow the same coursework as suggested by the Higher Education Council. Since, Turkish teacher education system is centralized and the Higher Education Council determines the structure and the content of all teacher education programs. However, the practical content of the courses may vary according to experience, research interests and initiatives of faculty staff in education faculties.

The current programs of faculty of education are resulted from reform efforts in teacher education programs throughout the country in 1998. The final program of elementary mathematics teacher education program was revised based on these reform efforts. As a result of field experiences, technology usage, methods of teaching mathematics courses became more important and are emphasized in the latest version of programs. According to current teacher education program, pre-service teachers are required to complete 153-credit coursework to become elementary mathematics teachers (See Appendix F). This coursework contains courses related to general pedagogical knowledge: including introduction to education, learning and development, classroom management, guidance; courses related to subject matter knowledge: including calculus and finite mathematics courses; and courses related to pedagogical content knowledge: including methods of teaching mathematics, school experience, and practice teaching. Pre-service elementary mathematics teachers are required to complete 15 - credit coursework for general pedagogical knowledge courses, 32- credit coursework for subject matter knowledge (advance mathematics) related courses, and 14 - credit coursework for pedagogical content knowledge courses. The courses related to general pedagogical knowledge spread homogenously throughout all grades. The subject matter courses are clustered in the first two years of the program, whereas courses related to pedagogical content knowledge are clustered towards the end of the program. In the third year of the program pre-service teachers are required to complete two methods of teaching mathematics courses, and in the fourth year of the program there are two field experience courses. Although there was some improvement on field experience courses in 1998 reform, pre- service teachers are dealing with observation tasks mostly. They have no opportunity to be actively engaged in teaching experience.

Although centralized programs exist for all faculty of education coursework, there are differences between the intensity of content and process of courses in different faculties. The content and the intensity of courses may also vary according to experience, research interests and initiatives of faculty staff in education faculties.

To sum up, the content and structure of the courses related to pedagogical content knowledge may have significant influence on the quality of teacher education system as well as the results of this study.

3.7 Administration Process of the Study

There were four main rounds of this study: Round One – item development and pilot testing; Round Two and Three– field-testing; and Round Four – validation. Accordingly, the data were collected at four different time intervals through different methods (i.e. qualitative and quantitative).Participants had been recruited from the departments of elementary mathematics education. Overview of the research design, major characteristics of the participants, instruments of the study, data collection and analysis procedures, and validation issues will explained in this section.

Table 3.3The summary of the test administration process

Round	*Time Interval*	*Participants*	*Purpose*	*Instruments*	*Type*	*Analysis*
Field Testing I	September 2010 to December 2010	44 PMTs 27 PMTs	Item Development and Pilot Testing	29 item-version of test Test Semi-structured Interviews	Quan. &Qual.Data Collection	Thematic analysis
Field Testing II	February 2011 to May 2011	1010 PMTs	Revision	16 item-version of two separate Tests	Quantitative Data Collection	Item Analysis Rasch Analysis
Field Testing III	September 2011 to October 2011	99 PMTs	Revision	20 item-version of Test	Quantitative Data Collection	Item Analysis Rasch Analysis
Field Testing IV	February 2012 - April 2012	167 PMTs	Validation	15 item-version of Test	Quantitative Data Collection	Item Analysis Rasch Analysis

3.8 Administration of Field Testing I

The purpose of the first administration was to understand how items worked. So both content analysis and qualitative interviews were conducted in Fall Semester 2010. Results of this implementation provided an opportunity not only to obtain in-depth feedback about the items, but also to understand how participants interpreted items and reasoned through the distracters. In this study, researcher obtained data from the pre-service teachers at their fourth year, translated and interpreted into meaningful information. Also, the researcher made the final decision regarding modifications of designed instrument.

3.8.1 Demographic Information of Participants in Field Testing I

The initial step of test development was to determine how items worked. In order to conduct an in-depth analysis convenient sampling was used at the first step of the study. The initial version of the instrument was administered to 44 fourth year participants at a state university in the fall semester of 2010. There were total 44 participants in the seventh semester of elementary mathematics teacher education program. There were 32 female, and 12 male pre-service teachers. They all completed the program requirements until that time. The minimum and maximum scores of CPG are 2.03 and 3.50 respectively (Median=2.69).

For the interviews, again, criterion-based or purposeful sampling techniques were used (Fraenkel & Wallen, 2006).The participants were selected according to following criteria: (a) accessibility, (b) their ability on expressing themselves to get in depth information, and (c) the differences in their perspectives in their open-ended responses. The data were collected through semi-structured interviews. There were 27 participants (10 male, 17 female participants) for the interview.

3.8.2 Implementation Procedures of Field Testing I

In this round of study two types of data were collected. The first one was administration of test itself. The second one was interviews with pre-service mathematics teachers.

For the initial version of the test there were 29 items presented along with five possible responses and the option to fill in an additional field (See Appendix A). The researcher collected data herself and participants were asked to answer the items and to comment on two topics. First, they were asked to explain their approach to answer item such as; why they chose certain answers and how they eliminated others. Second, participants were asked to point out any areas of confusion related to the content of the item and to express their recommendations (if they had).

The administration of test took about 65-70 minutes long for participants to complete the initial version of test. All pre-service teachers agreed to participate in a study voluntarily, and all of the participants were encouraged to give their best effort on the items. These results were used to revise distracters as well as stems of multiple-choice items. Illustrations, wording and content of some items were revised based on content analysis of open-ended responses. Based on the frequency analysis and cross-tabulation statistics distracters those did not work properly were either rewritten or replaced. In depth analysis of these findings will be explained in Chapter IV.

After the administration of the test, interviews with pre-service teachers was carried out to further understand how participants interpreted and reasoned through the items in order to obtain in-depth feedback about the items. The main purpose of the interviews was to check whether audiences could correctly interpret the items or whether there existed any other structural problems. Semi-structured interviews were performed using an interview protocol (See Appendix C). Findings of these interviews were used to clarify items as well as to revise the content and structure of distracters.

There were 27 participants interviewed on their responses about 40-60 minutes time interval. There were 10 male, 17 female participants. Before the interview starts, each interviewee was given their test booklets and list of the interview questions (See Appendix C). Then, pre-service teachers were asked to look at and remember the items about 10-15 minutes time interval with the help of interview questions.

During their review, they were asked to evaluate readability and understandability of wording, quality of illustrations as well as give-away characteristics of distracters, specifically focusing on any structure, which unintentionally helps test takers to answer the items correctly or prevent them to think in a correct way. Moreover, according to interview questions participants were asked to comment on how and why they chose certain answers. After they went on items for a while, participants were asked to think aloud on their pre-determined items. The main purpose of asking interpretations of items and their solution strategies was to check whether items were functioning as intended. Recorded interview sessions were transcribed and organized. The transcripts were utilized to better understand the findings from the initial administration. Analysis of transcripts provided an opportunity to get in-depth information about how items were interpreted by the test takers as well as the structure of items and distracters. In depth analysis of these findings will be explained in Chapter IV. Based on these interviews, the items were revised for clarity and distracters were improved. After required revisions the next version of the test was administered in Spring Semester 2010-2011.

3.8.3 Data Analysis of Field Testing I

At this step of test development qualitative data analyses were conducted:

Qualitative data analysis included: a) the content analysis of open-ended responses, b) analysis of interview results.

3.8.3.1 Content Analysis of Open- ended Responses

In the first phase, all multiple-choice items were coded according to the rubric given in Table 3.5, and frequency analysis was conducted to see results of how keyed items and distracters worked. This analysis was conducted in hopes of eliciting ideas that may have been overlooked during item construction. Rubric up to two points was prepared; comments or claims in open-ended field were coded according to the rubric. In fact, this rubric was prepared to determine the malfunctioning distracters, in such a way that any distracter having 2 point might be an indicator of overlooked idea related to the distracter. On the other hand, any keyed response has 0 or 1 point suggested to investigate in detail.

Table 3.4Rubric for assessment of open-ended comments

Score	*Content*
2 points	▪Any comment, claim, or strategy, which is logically and mathematically correct - (even if computation errors exist)
1 point	▪Any comment or claim, which discriminates the problem in the given case, but somehow fail to determine true answer. ▪Any comment, claim, or strategy that suggests a mathematical solution but is not completely true.
0 point	▪No answer. ▪Completely wrong comment or claim.

Cross tabulation with distracters and open-ended explanations gives idea about functionality of distracters. Results of this analysis helped to revise distracters in detail. For example, if there was logically and mathematically correct explanation to any alternative previously determined as correct answer, this observation forced to make necessary revisions related to either stem or distracters.

3.8.3.2 Analysis of Interview Findings

In the second phase, interview results were transcribed. The transcripts were divided into episodes item by item. Those comments on specific items were categorized into two main themes (See Table 3.6). The first one was necessary revisions to improve the understandability of item, and the second one was a set of challenges, which resulted from the interviewee’s background knowledge. The first category was divided into three subcategories such as revisions necessary for visual of the item (visual), revisions necessary for linguistic structure of the item (linguistic), revisions necessary for mathematical content of the item (mathematical content). Each category referred the necessary revision for improvement of either stem or distracters. The second one was coded as challenge resulted from the interviewee's background knowledge (background knowledge). These themes help to determine necessary improvements to revise items analytically.

In order to be more specific, this process was explained for the case of LS2. After transcription of the interview data, statements of all participants related to LS2 were combined. Then, all comments were categorized according to four general themes stated above. The researcher then refined and reformulated initially proposed items using the patterns that emerged from the participants' responses. Three different categories of item modifications suggested by participants; a) visual, b) linguistic, and c) mathematical content (see Table 3.6). Comments on the LS2 in each theme were gone over and necessary improvement was done.

Table 3.5Rubric for assessment of interview findings

Themes	*Findings*	*Revision*
Visual	"Red line on the visuals are not clear in photocopied sheets" "Not clear enough how balls are positioned inside"	▪Red lines turned into black ▪Visible balls put inside
Linguistic	No comment	▪Based on other comments in the test, statement of item 2 also revised and the stem of the item became shorter.
Mathematical Content	"No information about the sizes of balls as well as cylinders. Are they identical or not?"	▪Term of "identical" was used to describe both cylinders and balls.
Background knowledge	"No value of π. We cannot be sure the result of comparison. Depending on the value of π, it is either equal or one is longer than other".	▪ No revision

An example of how interview findings affect the structure on items was given in Table 3.6. For the initial version of LS2 as in the Figure 3.2, participants commented the weaknesses of visual of item. The red color of security tapes in master copy resulted in fuzziness in further copies, since further copies duplicated in black and white. As a result, it was necessary to change the color of security tapes in master copy. Another comment was about the participants' difficulty in imagination about how tennis balls organized inside the packages, whether there exists space around balls or not.

Balls were assumed to be close-fit to packages and there was no extra space around balls during item writing process. From this perspective, the issue was quite clear for researchers, and accordingly it was overlooked until implementation of test and interviews. For later implementations, packages were become transparent, and balls became visible inside. For this item interviewees did not make any comment on the linguistic structure of the item. But in order to shorten duration of implementation, researchers revised the stem of the item as seen Figure 3.3.

Another important comment was about the mathematical content of the item. Similar to researchers' assumption on the positions of balls inside packages, sizes of both cylinders and balls were assumed as identical. But there were no information stated in the stem of LS2. But participants wanted to be sure about the sizes of cylinders and balls, asked to confirm the identicalness during not only administration of the test but also interviews. Dependently, the term of "identical" was used to describe both cylinders and balls in item stem. Finally some interviews and some open- ended comments stated that they might not be sure about the result of this comparison.They stated that they sometimes used to assume the value of π as merely 3 to simplify the complicated computations. If they had assumed the value of π as 3, lengths of these two security tapes equal, on the other hand if they had used the value π as it is, the length of security tape in Case 2 (stated as Figure II in the item stem) would be longer than the Case 1 (stated as Figure I in the item stem). So, they confused with the result of comparison and there was no clear answer for this case. These statements were directly related to the participants' background information, not related to the structure of item. Thus there was no revision made for these comments.

Example: The development of Item_2 based on interview findings

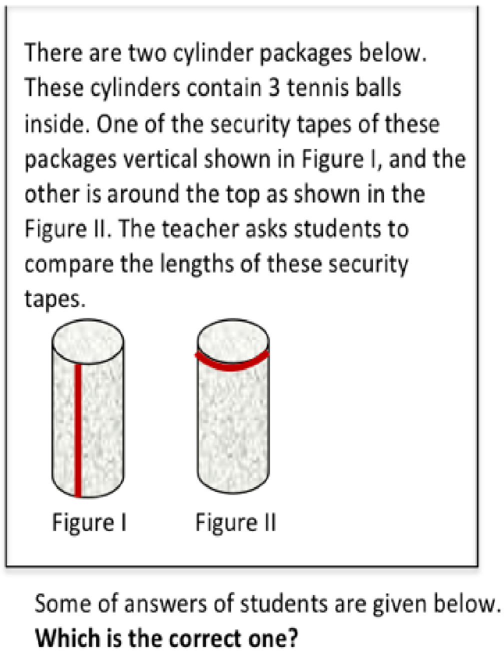


Figure 3.2 Initial version of LS2

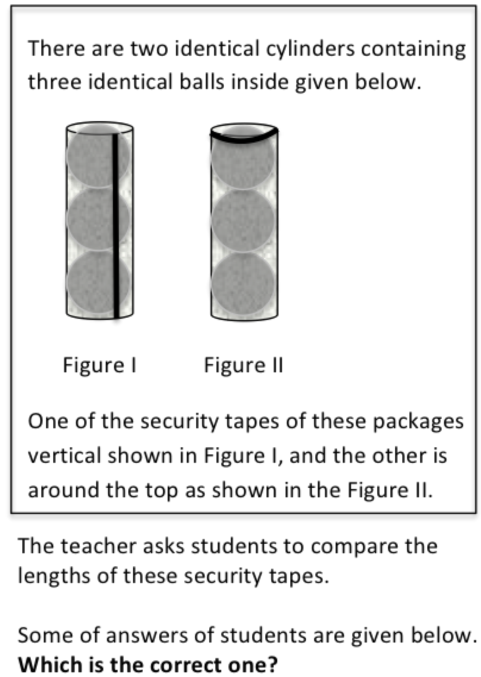


Figure 3.3 Final version of LS2

As exemplified for LS2, qualitative data analysis was conducted in this manner and data analysis was carried out for all items discussed in interviews.

3.9 Administration of Field Testing II

The purpose of the second administration was to understand how items worked from the perspectives of Item Response Theory and Classical Test Theory. A quantitative research design -survey method-was utilized to gather data for the study in Spring Semester of 2010-2011. Both Rasch analysis and Item analysis were conducted in order to see how items functioned.

3.9.1 Population and Participants of Field Testing II

Participants have been recruited from elementary mathematics education departments from almost all districts of Turkey. In this step, random sampling was used to determine the sample. All names of elementary mathematics teacher education departments which had 4th grade pre-service teachers were enrolled in the Microsoft Office Excel program. By the help of random function of the program the sample of Field Testing II was determined.There were totally 1010 pre-service teachers from 17 universities. The summary of demographic information of participant was given in Table 3.6.

Table 3.6Frequency distribution of demographic information of participants (n=1010)

	Frequency (f)	Percentage (%)
Gender		
Male	346	34.3
Female	649	64.3
Missing	15	1.5
Total	1010	100
Type of Program		
Regular Classes	615	60.9
Night Classes	393	38.9
Missing	2	0.5
Total	1010	100
Teaching Experience (hour)		
No experience	305	30.2
Less than 10	115	11.4
11-20	130	12.9
21-50	187	18.5
51-100	114	11.3
More than 101	159	15.7
Total	1010	100

Types of program that participants enrolled also summarized in Table 3.6. According to Table 3.6, more than half (60.9 %) of the total participants were enrolled in the regular program, whereas 38.9 % of participants followed the night classes teaching program. Participants were also asked their own teaching experience. They were advised to answer this item considering all teaching activities those they had actively participated in, such as teaching practices, voluntary teachings, private lessons etc, except for classroom observations. Information related to teaching experience was displayed in Table 3.7. As seen in Table 3.7, about one third of the participants (30.2 %) had no teaching experience; the rest of them had varying levels of teaching experience. Only about one sixth (15.7%) of participants had more than 101 hours teaching experience.

Moreover, types of high school that students graduated were taken into consideration. Information related to school types, which they graduated from, was summarized in Table 3.8. According to Table 3.8, almost half of participants have had background of teaching education since their high school education (47.3%), since they had graduated from Anatolian Teaching High School.

Table 3.7Frequency distribution of the participants according to school types they graduated

School Type	Frequency (f)	Percent (%)
General High School	131	13.0
Anatolian High School	252	25.0
Anatolian Teaching High School	478	47.3
General High School with Intensive Foreign Language Program	93	9.2
System Missing	56	5.5
Total	1010	100

Almost all participants completed the program requirements until data collection procedure started. This means that they completed courses related to mathematical content such as Calculus I and II, and Geometry as well as courses related to teaching such as Teaching Methods I and II, and School Experience. Summary of related information was given in Table 3.9.

Table 3.8Frequency distribution of participants who would graduate at the end of the semester which data was collected (n=1010)

	Frequency (f)		
Course	Complete	Did not complete	System missing
Teaching Methods I	1004	4	2
Teaching Methods II	984	14	12
Geometry	1004	2	4
Calculus I	1006	0	4
Calculus II	1005	1	4

3.9.2 Data Collection of Field Testing II

Items on each previous version of the test were analyzed the Spring 2010 version. A few items had undergone minor revisions and had been unchanged. Three bad working items were eliminated. Required changes related to rest of items were made for further implementations.

The initial administration revealed that 29-item test was too long and too difficult to complete the test within the allotted time limit. Based on this feedback it was decided to split items into two parallel tests in order to shorten the duration of implementation. These items were included in the data set but were divided into their two versions, for example Test 1and Test 2. The data included in the master set was shown in Table 3.4 and table of specifications of these tests are given in Table 3.10. The common items serve as “anchor items” and item parameters are estimated simultaneously.

After constructing two separate tests just to see whether there existed any critical difference between tests, these two tests were piloted in one state university. Initial item analysis results indicated that there was no significant difference between item statistics of two booklets. Based on this observation, it was decided to administer these tests as they were. The next version of the instrument involved two sets of items (each contained 16 items with 6 anchor items).

Table 3.10 Table of specification of two tests

	SCK		PCK	
	Test 1	Test 2	Test 1	Test 2
The concept of Measurement	MS20, MS8	MS20, MS8		
Length Measurement	LS2, LS12	LS6, LS11, LS16, LAS25	LAP14,	LAP14, LAP7, LP29
Area Measurement	AS19, AP17	AP13, LAS25	AP3, AP15, AP4, LAP14, AP5, AVP16, AP10	LAP14, AP21, AVP26, LAP7, AP4
Volume Measurement	VS27, VS18, VS24	VS9, VS27	AVP16	VP1, AVP26

Frequency Distribution of Booklets

Two different booklets were assigned to participants randomly during the implementation. Frequency distribution of the booklets was displayed in Table 3.12. As seen in Table 3.11, the results indicated that the booklets were almost evenly split among participants.

Table 3.11 Frequency distribution of booklets

Booklet	Frequency (f)	Percentage (%)
Test 1	504	49.9
Test 2	506	50.1
Total	1010	100

Moreover, the researcher administered all tests. During the administration, in each classroom equal-number of the booklets was assigned to participants randomly. Moreover in order to eliminate extraneous variables those possibly affect the results of the study; such as cheating any other manipulations, were taken under control. Since the booklets were assigned randomly, other demographic information of participants -such as gender, program type they attended, teaching experience, graduation status, school type they graduated from- were also distributed almost in the same manner among participants who took different tests.

3.9.3 Data Analysis of Field Testing II

The instruments contained 16 items and participants were given approximately 35 minutes to complete it. The data obtained from the administration were tabulated and entered into a computer spreadsheet. The scores from each subtest were entered into a separate spreadsheet. The data in each spreadsheet were then visually inspected for invalid entries. After the data were "cleaned" each spreadsheet was transformed into the correct file format (i.e., .dat) for both ITEMAN Version 3.6 by Assessment Systems Corporation 1994, and BIGSTEPS (Wright & Linacre, 1991) -computer programs for Item analysis and for Rasch analysis respectively.

First, in order to guide the test development process Rasch analysis was conducted. Once the data set had been created, the Rasch analysis was carried out using the analysis software BIGSTEPS. The data were modeled with a Rasch model, which estimates two parameters for the data: one is for person trait (θ) and item difficulty (β).

After Rasch analysis, Item Analysis was conducted with the same data for the perspective of Classical Test Theory. ITEMAN Version 3.6 was used for this purpose. For each item, the difficulty, discrimination index, correlation with the total score, and alpha-if-item-deleted values were determined. Misfunctioning items were either eliminated or revised based on the item analysis results.

3.9.3.1 Rasch Model

Among all IRT models Rasch model has the fewest variables, one parameter for trait level (θ) and one parameter for item difficulty (β). For this reason the Rasch model is called as one-parameter model. At this point one-parameter refers "one" parameter for the difference between person position and item difficulty. In fact, model actually estimates two parameters: person trait and the item difficulty. Parameters of person trait(θ) and item difficulty (β) are used to calculate the probability of person (i) succeeding on item (j) (Wright, 1977).

3.9.3.2 Assumptions of Rasch Analysis

There are three main assumptions for Rasch model similar to other commonly employed IRT models. The first one is the sample size, different from other models of IRT, smaller sample size might be enough to produce conclusions for Rasch model. The second one is unidimensionality. The other assumption of Rasch model is the local independence, which is parallel to the second assumption.

Sample Size

Although there are various standpoints about the proper sample size for parameter estimations (Hambleton, 1989), Rasch analysis, however, requires fewer samples than the other IRT models. For the Rasch analysis, the minimum number of sample size recommended for 20-item test is about 200 examinees (Wright & Stone, 1979).

Unidimensionality

Rasch analysis assumes the presence of a dominant ability or trait that influences test performance- which is called unidimensionality (Hambleton et al., 1991). In other words, unidimensionality refers that there exist a single latent trait variable to explain the variability of observed score as well as assumption for the test development in classical test theory.

Locally Independent Items

The other assumption of Rasch model is the local independence, which is parallel to the unidimensionality assumption and suggests that there is no correlation between test items when person's trait level is controlled (Hambleton et al., 1991). This simply means that any item in the test should not affect the examinee's response of any other item. This, therefore, requires that "the content of one item must not provide any clues to the answer to another item" (Hambleton & Swaminathan, 1985, p. 23).

3.9.3.3 Test and Item Analysis within Rasch Model

Rasch Model is basically based on the idea that the results of unidimensional measurement (examination of only one human attribute at a time) can be ranked along a vertical line and this vertical line provides an opportunity for in-depth investigation of data to researchers (Bond & Fox, 2007).

Hierarchy of items

One other useful representations of a Rasch analysis is the distribution map of items and persons. Almost all software tools for The Rasch Analysis can graph person position with item position. A distribution map of items and persons includes several key features including central line marked out in logits typically ranging from -2 to +2), which determine the relationship of the construct to the probability of response. The left hand side of the distribution map includes the locations of respondents- indicated by a " # ", while the right hand side is the item locations in terms of item difficulty on the logit scale (Bond & Fox, 2001).

Placement of items and persons on a common scale permits evaluation of test function relative to the sample. This item-person graph is useful in three ways: (1) to determine the extent to which item positions match person positions (appropriate, too easy or too hard), (2) to detect the gaps in the measure, which suggests where items might be added, (3) to assess the validity of the measure by reviewing the item order.

The distribution map of the items and persons gives idea about whether the calibration process provided useful information. This table summarized the information about the ability of examinees and difficulty level of items. The most able persons and the most difficult items lined up upwards along the vertical line.

Fit Analyses

In order to benefit the advantages of Item Response Theory, the goodness of fit should be provided. One of the emprical ways that this criterion can be assessed is by using fit indices. Fit statistics provide the indices of fit of the data to the model and appropriateness of the measure (Bond & Fox, 2001).

The statistics of items and persons are estimated in terms of logit scale with some degree of error. There are two types of indicators of misfit in Rasch Model. To begin with, the first misfit index is overfit (muted), which is mean square value lower than 1.0, in other words, a negative standardized fit. Overfit implies little variation in the response pattern, perhaps indicating the presence of redundant items. The second misfit index is underfit (noise) is a mean square >1.2 and standardized fit >2.0 and suggests unusual and/or inappropriate response patterns. These indices can be used to identify and sometimes correct a measurement disturbance (Linacre, 2007).

Person fit indices are indicator of how individuals respond to items. Linacre (2007) exemplifies these cases such that noisy infit and outfit indices may give clues about their inconsistency of responses. For example, noisy outfit index of a high ability person might be an indicator of inattentiveness, confusion, carelessness, rush or sleeping behaviors. Similarly, noisy outfit index of a low ability person might be an indicator of guessing or any item targeting special knowledge. On the other hand, muted infit index of low ability person may indicate caution or plodding.

Similar to person fit indices, item fit indices are indicator of how items function logically and provide a continuum useful for all respondents. For example, noisy outfit index of a hard item might be an indicator of ambiguity, negative wording, or misleading options. On the other hand, muted infit index of an item may indicate similar items in the test, correlation with other items in test (Linacre, 2007)

Person reliability and separation:

The other series of analyses address another Rasch measurement principle the important criterion of reliability. In Rasch analysis, reliability is calculated by the item separation index and the person separation index. Separation indices are useful for the assessment of test spread across the trait continuum. Both items and persons spread along the continuum in standard error units. A value of 1.00 represents high separation ability, in which errors are low and item difficulties and students' measures are well separated along the scale (Wright & Masters, 1981). Person and item separation and reliability of separation indices can be thought of as the number of levels into which the sample of items and persons can be separated. For a test development it is very useful such that if separation exceeds 1.00, with higher values of separation representing greater spread of items and persons along a continuum. Higher separation index implies higher variance in person or item position, and then it yields higher reliability. In Rasch analysis, Reliability of person separation is conceptually equivalent to Cronbach's alpha. However, the formulas of these two are, in fact, different. Lower values of separation indicate redundancy in the items and less variability of persons on the trait.

Linacre (2007) explains that there are several factor those affect the value of the reliability, the first one is the variability in the ability of sample, wider ability range implies higher person ability. The second one is the length of the test. Similar to classical test theory longer tests imply higher person reliability. Sample-item targeting is another factor that has an effect on the reliability value. Too hard and too easy tests provide lower reliability values. Finally, higher number of categories among items provides higher reliability.

Person separation index is an indicator of the spread in person measures. This index indicates the number of distinct levels into which the sample of persons can be classified. This degree of separation indicates that the difference must be due to the differences in the magnitudes of the person's underlying attribute (Bode & Wright, 1999). A variable is useful only if persons differ in the extent to which they possess the trait. The larger the item separation, the wider the range of the attribute defined by the set of items.

Item separation index is an indicator of the spread in item difficulties. This index indicates the number of distinct levels into which the sample of items can be classified. This degree of separation indicates that the difference must be due to the differences in the magnitudes of the items' difficulty level (Bode & Wright, 1999). The larger the person separation, the wider the range of the attribute defined by the set of persons. A variable is useful only if item differ in the extent to its difficulty. In order to determine the extent to which the test distinguished among persons with different levels of functioning and participation (with separation index criterion set at ≥ 2.0, and reliability at ≥ 0.50).

3.9.3.4 Classical Test Theory

CTT has traditionally been used to validate the psychometric integrity of new measures and it has comprehensive literature. This section will outline the basic concepts of CTT as well as highlights its strengths and limitations. The information about basic model of CTT stated below is common in almost all psychological testing sources; as a result, this information was summarized without any specific reference.

Classical test theory provides a model about test scores that introduces three concepts- *test score (*often called the *observed score*),*true score,* and *error score*. Within that theoretical framework, models of various forms have been formulated. For example, in what is often referred to as the "classical test model" a simple linear model is postulated linking the observable test score (X) to the sum of two unobservable (or often latent) variables, true score (T), and error score (E), that is,

$$\text{Test score (X)} = \text{True score (T)} + \text{Error score (E)}$$

From this linear formulation, classical test theory assumes that the observed score is the sum of the true score and some error, which includes the ability of the examinees or the difficulty of items. In addition, true score of a person could be calculated hypothetically by the mean of scores of the same test if a person had an infinite number of the same test.

3.9.3.5 Assumptions of Classical Test Theory

One of the well-organized summaries of assumptions for CTT can be found in the related chapter of Gruijter and Kamp (2008, p.13). They listed basic assumption of CTT as following:

(a) Classical test theory assumes that error scores are constant, i.e. true scores and error scores are uncorrelated. As a result, the error score will neither decrease nor increase due to a change in the true score.

(b) The average error score from the examinee population is zero in the long run.

(c) Error scores on parallel tests are uncorrelated. This implies that parallel tests are two tests that yield the same true score and the same variances of error score from the perspective of classical test theory.

(d) Repeated administrations of a test results in a value of the observed score exactly equal to that of the true score.

3.9.3.6 Test and Item Analysis within Classical Test Theory

In order to determine how items work in relation to the other items and whole test as well, there exist several tools to get useful information from the perspective of CTT.

Descriptive Statistics: In order to see the overall picture about results of the test, it is suggested that descriptive statistics be conducted. These statistics consist of especially mean and variance. The values of mean and variance give the general idea about the structure of the whole test and test items individually as well.

Item Difficulty: The item difficulty is a measure of the percentage or the proportion of participants who answer the item correctly (Crocker & Algina, 1986). For dichotomously scored items (multiple choice items) item difficulty is calculated by the percentages of the true choice and the item difficulty (p) ranges from 0 to 1. If items are harder then as expected the fewer people will reach a correct answer and items will have low item difficulty, whereas easier items have high item difficulty. There is not a common agreement on the ideal value for item difficulty. However, items that are extremely easy or extremely difficult decrease the total variance of the test. For dichotomously scored items, mean score represents this proportion and total test score variance is assumed to be maximized when the $p = .50$ (Crocker& Algina, 1986). It was decided that a wide range of item difficulties ($0.2 < p < 0.8$) was desirable. Items with outlying p-values were flagged for revision or elimination pending further administration of test. Item difficulty scores of items may indicator of poorly written items. For example, a high p-value may also be an indicator of a poorly written item in which the correct answer was somehow obvious to test takers based on factors unrelated to the conceptual content of the item. On the other hand, a low p-value does not necessarily indicate a malfunctioning item. A good item can sometimes be answered incorrectly by most of students because it addresses a particular misconception or reasoning difficulty.

Discrimination Index: The discrimination index is a measure of how well an item separates high scorers from low scorers of the test. For dichotomously scored items (multiple choice items) discrimination index for an item is calculated by using upper 27% and lower 27% of the examinees and item discrimination (D) ranges from 0 to 1. The discrimination index gets its maximum value of 1 if every participant in the upper group answers the question correctly whereas every participant in the lower group does not. Items with a large, positive discrimination index are assumed to be good, which gives evidence that the items are measuring the same construct and contributes to the reliability of the test. Items with a low discrimination index may need to be rewritten or reconsidered. If the items are hard or too easy, most of the participants may miss the item; thus, discrimination index will be low.

Criteria for the discrimination index:

$D \geq .40 \rightarrow$ very good items,

$39 \geq D \geq .30 \rightarrow$ reasonably good but possibly needs to be revised,

$.29 \geq D \geq .20 \rightarrow$ marginal items, usually needs to be improved,

$.19 \geq D \rightarrow$ poor items, to be rejected or improved by revision.

(Ebel, 1979, p.267).

Correlational Indices of Item Discrimination: The correlation index is a measure of how well an item correlates each item score with the whole test score. A positive value of the point biserial coefficient indicates a positive correlation between the item score and the overall test score. Likewise, a negative point biserial coefficient signifies that high scorers are answering the item incorrectly more frequently than low scorers and may signify a problem with the question. In achievement testing, if an item is functioning good enough, then high scorers have a high probability of answering this item correctly, whereas low scoring examinees have a low probability of answering it. So, the following is expected from a discriminating item:

- key option is selected by high achievers,
- distracters of an item are selected by low achievers,
- high achievers do not prefer to omit the item (Crocker& Algina, 1986).

For dichotomously scored items (multiple choice items) correlational discrimination index for an item is calculated by calculating the correlation between the item score and the total score for the rest of the items. There are two types of calculating correlational discrimination index; point biserial and biserial correlation coefficients. Point biserial and biserial correlation coefficients (r) typically range from zero to 0.4. Especially, for the instruments that have a small number of items point biserial and biserial values above 0.2 are considered good (Nunnally, 1967). The higher the item correlations with the total test score are the more contribution there is on the test score and the more reliable the test is. However, items with correlation indices lower than 0.30 may be considered for retention if the point biserial is significantly greater than 0.00 (Crocker & Algina 1986; Ebel 1965). For determined sample size, the minimum critical value of the point biserial coefficient is set at two standard deviations above 0.00 ($0.00 + 2\sigma_p$).

Overall Alpha Rank: The coefficient alpha (α) is a measure commonly used to estimate the reliability of an instrument as a whole. Reliability evidence of the scale is assessed by internal consistency reliability (i.e., Cronbach's alpha). Cronbach's alpha is a whole test measure; it can only be computed for single administrations of the instrument. Coefficient alpha is, also, sample dependent and can vary depending upon the characteristics of the sample. However, this sometimes results in difficult to interpret it.

3.10 Administration of Field Testing III

According to item analysis statistics of second round 20 items were selected for third administration. For third round of the field-testing, relatively good functioning 20 items were used for data collection. The purpose of the third administration was to understand how selected items worked together. Similar to Field Testing II, a quantitative survey research design was utilized to gather the data for the study in Fall Semester of 2011.Both Item analysis and Rasch analysis were conducted in order to see how items were functioning and to obtain in-depth feedback about the items.

3.10.1 Participants of Field Testing III

The second version of the instrument was administered to 99 pre-service elementary mathematics teachers enrolled in two state universities in Ankara voluntarily. In total 99 pre-service teachers; 79 of them were female (79.8%), 19 of them male (19.2%), and 1 system missing, from 2 state universities. This was a sample of convenience. Majority of participants, who took the tests, completed the program requirements until that time. This means that they completed courses related to mathematical content such as Calculus I and II, and Geometry as well as courses related to teaching such as Teaching Methods I and II, and School Experience.

3.10.2 Data Collection of Field Testing III

Based on both results of Item Analysis and Rasch Analysis of previous administration, poor functioning 6 items were eliminated. Some remaining 20 items were revised and made small changes. Those 20 items were administered in the fall semester of 2011.

3.10.3 Data Analysis of Field Testing III

The researcher followed the same procedure of Field Testing II for data analysis. Both item analyses and Rasch analyses were conducted.

3.11 Administration of Field Testing IV

The purpose of the fourth administration was conducted for validation study. Similar to Field Testing II and Field Testing III a quantitative survey research design was utilized to gather data for the study in Spring Semester of 2011. Item analysis, Rasch analysis and Factor analysis were conducted to get an in-depth feedback about items and test in general. According to item analysis statistics of third round 15 items were selected for fourth and last administration. For fourth round of the field-testing, good functioning 15 items were used for data collection.

3.11.1 Population and Participants of Field Testing IV

There were total 168 participants voluntarily participating in this implementation. Participants have been recruited from elementary mathematics education departments from three state universities and one private university in Ankara. There were totally 168 pre-service teachers; 146 of them were female (86.9%), 22 of them male (13.1%), and 3 universities in Ankara. The descriptive results indicated that similar to other rounds. Majority of Round III sample was female (86.9%). But for this round, the number of female participants was almost five times the number of male participants.Different from other three rounds, in this round there were also 3rd grade pre-service teachers. The main reason why the third grade pre-service teachers had been included into this round was to compare the mean scores between fourth and third grade of pre-service teachers. There were totally 99 (58.9%) third grade pre-service teachers enrolled in this round, on the other hand, there were 69 (41.1%) fourth grade pre-service teachers.

3.11.2 Data Collection of Field Testing IV

Based on both Item Analysis Results and Rasch Analysis Results poor functioning 5 items on the third iteration were eliminated. Remaining 15 items were revised. Those 15 items were administered in the Spring Semester of 2011-2012.

3.11.3 Data Analysis of Field Testing IV

The researcher followed the same procedure for data preparation, item analyses and Rasch analyses, and consideration for removal of participants or items.

3.12 Quantitative Validityof Test

Validity is an important concept in the test development process. Messick (1989) defines validity as "the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores" (p. 13). Validity is a property of test scores and the inferences and decisions that are made based on them. There are different types of validity (e.g., content validity, criterion-related validity, predictive validity, concurrent validity) defined in literature. But, Loevinger (1957) and Messick (1989) both argue that construct validity encompasses all other types of validity. Based on this conceptualization, they point out three components of construct validity: *substantive, structural* and *external validity* of measures.

Substantive validity comprises the conceptualization of the domain and development of an initial item pool for measures. The substantive validity goes beyond making this claim based on the fact that the items are really represent intended construct. This validity can be derived from a domain space clearly specified in advance and judged by experts to be representatives of the area. Statistical analysis should confirm that the items should function consistent with the construct, as well as consistent with each other. Analysis of items should include individual analysis of keyed answer and distracters.

Face and content validity were determined by consulting in-service and pre-service teachers, mathematics educator, test development expert and test developer. In order to ensure the substantive validity of the test, four in-service teachers evaluated the reality and appropriateness of cases in the item stems with real classroom contexts. Besides in-service teachers, three experts were involved in evaluating the structure of test. Content validity was also addressed by using the expert panel to ascertain adequate topic coverage. One expert who is specialist in mathematics education not only investigated content of items but also checked with respect to the categories of intended domain in the table of specification. One test development specialist analyzed the content and structure of items in terms of test development principles. One test developer investigated the wording and language of items for readability and understandability. Before implementing items, 7 participants asked to complete the test and make comments and criticisms for face validity of test. Further, the issues of content and construct validity were addressed by the extensive literature review. To sum up, every attempt to provide evidence in test development steps also supported the substantive validity of the test.

Structural validity of the measurement is complementary to the substantive validity and mainly encloses the item selection and psychometric evaluation of the test (Messick, 1989). According to his definition, the substantive validity relies heavily on the analysis of test scores. Structural validity issues of test were addressed by conducting individual analysis of keyed and distracter responses. Item analysis included item analysis including calculating item difficulty, item discrimination, and correlation coefficients. The values of these coefficients provided evidence whether it was a good item in terms of general objective test theory. Each step explained in Item analyses and Rasch analyses provided evidence to support the structural validity of the test. Classical Test Theory provided an opportunity to analyze cumulative scoring; on the other hand, Rasch Model provided an opportunity to analyze items as individual responses.

Finally, the *external validity* of how the test interacts with other measures (both test and non-test behaviors) as predicted by construct theory (Messick, 1989). Thus, external validity of the measure depends on the quality of initial steps of test development. Validation work begins to assess; using non-psychometric methods, whether the items tap the intended construct. Briefly, the validity of each step in test development will ensure the whole validity of the measure.

3.13 Qualitative Trustworthiness of Test

The study involved both qualitative and quantitative data collection. The discussion of trustworthiness was considered according to following categories stated in Lincoln and Guba (1985). Trustworthiness of qualitative data was addressed in terms of credibility, transformability, dependability, and confirmability. Credibility refers to the credible conceptual interpretation of qualitative data drawn from original responses of participants. Transferability refers the degree of transferring qualitative findings beyond the research context. Dependability refers the quality on the integration of data collection and data analysis. Finally, confirmability refers the degree of supportiveness of research findings by qualitative data. In order to ensure the trustworthiness of qualitative findings of this study, there were some steps used systematically as suggested in Creswell (2003) ; (a) triangulation of data during item development process, (b) member-checking, (c) using peer debriefing, (d) clarifying the bias of the researcher, (e) detailed description and reporting item development process.

The qualitative findings on how items were functioning were triangulated by item analysis results, content analysis of open-ended responses, and interview transcripts. In order to ensure the qualitative findings, the interpretations of the findings were checked and confirmed by some of the participants. One impartial colleague who was knowledgeable about not only mathematics education research but also content and context of the study critically review the implementation and evolution of test development steps. During interviews, researcher did not attempt to intervene the participants' expressions continuum, and the semi-structured interview protocol was also used to minimize the effect of researcher bias.

CHAPTER 4

RESULT

Overview of the research design, major characteristics of the participants, instruments of the study, data collection and analysis procedures, and validation issues were explained in the previous chapter. This chapter will provide an overview of the development procedures of the test from the beginning to the whole test. There were four main rounds of this study: Round One – item development and pilot testing; Rounds Two and Three– field-testing; and Round Four – validation. Accordingly, the data were collected at four different time intervals through different methods (i.e. qualitative and quantitative).Participants were recruited from the departments of elementary mathematics education from almost all districts of Turkey. This participatory study was conducted from the semesters of Fall 2010 to Spring 2012. Based on this design, in this chapter, qualitative findings and quantitative results of test development procedures and results of quantitative data analysis of survey data will be explained.

The process of test development was an iterative process: the content coverage and item functionality was investigated from the data collected from the Round 1; initial item properties was investigated by using CCT and IRT analyses on the data collected at Round 2 and Round 3. Many variations of items were analyzed to maximize the item and test properties maintaining the content coverage across the range of participants' abilities.

The description of the results of the study begins with qualitative and quantitative analysis, which framed the open- ended responses and interview findings. In the first section, both quantitative results, i.e. item and distracter analysis to understand how items worked, and qualitative findings, i.e. how PMTs interpreted items and reasoned through the distracters will be explained in detail. Moreover final decisions on the item selection process will be summarized at the end of the first section. In summary, a brief history of how the test was developed will be presented including topic selection, question development, and revision practices.

In the second section, results of quantitative data analysis of survey data will be presented. In other words results of how items worked will be presented considering the perspectives of both Item Response Theory and Classical Test Theory. The Rasch analyses and Item analyses of Round 2 and Round 3 will also be presented to determine a workable but reduced set of items in the second section. Along with these, analyses of dimensionality and fit analyses were performed to check the unidimensionality and fit assumptions. A presentation of the results of the field-testing of the items will be explained with the most emphasis on the answer of the Research Question based on the data from Round 2 and Round 3.After selecting the final set of items, lastly, results of fourth administrations will be presented respectively.

4.1 Development of the Test and Results of Field Testing I

The main purpose of the first administration was to understand how items worked. So both content analysis and qualitative interviews were performed with the purpose of conceiving how the items worked at this step. In this version of the test there were 29 items presented along with five possible responses and the open ended option to write additional comments (See Appendix A). PMTs were asked to comment on why they chose certain answers and how they eliminated others as well as to point out any areas of confusion. After the implementation of TKM-M test, semi-structured interviews with PMTs were carried out to further understand how pre-service teachers were interpreting and reasoning through the items and to obtain in-depth feedback about the items. Since the purpose of the step was to understand how items worked, quantitative analysis of test scores and reliability analyses were not conducted at this step.

The content analysis was conducted to analyze the responses stated in open-ended fields of items. This analysis was conducted with the purpose of eliciting ideas that may have been overlooked during item construction. Rubric up to two points was prepared for each item, and comments or claims in open-ended field were coded according to the rubric displayed in Table 3.6 in Chapter 3. Cross tabulation with the frequency of distracters and encodes scores of open-ended explanations gives idea about functionality of distracters. Results of this analysis helped to revise distracters in detail. For example, if there was logically and mathematically correct explanation to any alternative previously determined as incorrect answer, this observation forced to make necessary revisions related to either stem or distracters.

Besides analysis of open-ended responses, interviews with pre-service teachers were carried out to further understand how students interpreted and reasoned through the items and to obtain in-depth feedback about the items. The main purpose of the interviews was to understand how items worked. Semi-structured interviews were performed with an interview protocol (See Appendix D).There were 27 participants interviewed on their responses about 40-60 minutes time interval. There were 10 male, 17 female PMTs. Before starting interview, participants were given 10-15 minutes occasion to review items and recall how they had reasoned their answers through the given list of interview questions. After they went on items for a while, PMTs were asked to think aloud on how and why they chose certain answers. The nature of the items and how they functioned derived from the descriptive results were discussed in the context of interviews. This analysis provided an opportunity to get in-depth information about how items were interpreted by the test takers as well as the structure of items and functionality of distracters. Transcripts were utilized to better understand the findings from this initial administration. Those transcripts were divided into episodes item by item. Those data on specific items were coded and categorized into two main themes according to the rubric explained in Table 3.7 in Chapter 3. The rejected data related to each item of the instrument was also reviewed along with the reasoning for why these data was not taken into account.

Results of this implementation provided an opportunity not only to obtain in-depth feedback about the items, but also to understand how PMTs interpreted items and reasoned through the distracters. These results were used to revise distracters as well as stems of multiple-choice items. Some illustrations of items and content of some items were revised based on the statements of PMTs'. Distracters those did not work properly were either rewritten or replaced. How the analysis was conducted and the modifications were made were described in detail in Chapter 3.

4.1.1 Analysis of Open Ended Responses

Based on the initial-phase detailed analyses described in the previous section, the summary of content analyses of open-ended responses and related revisions of items was summarized in the Table 4.1. As seen in the table, there were two main categories related to responses: correct and incomplete/incorrect answers. For each item, the numbers of correct and incorrect answers were presented. Moreover there were mean scores of the open-ended responses were analized. Based on this table if there was logically and mathematically correct explanation to any alternative previously determined as incorrect answer, this observation forced to make necessary revisions related to either stem or distracters. Other revisions were made based on interview findings. These changes will be explained briefly in this section.

VP1 asked PMTs to diagnose the reason why student made mistake given in the stem regarding the relationship between the cylinder radius and volume, and asked them to suggest an activity to help such student overcome his difficulty (See Appendix A). Responses on alternatives had clear diagnosis of student's error given in the stem of the item. But some of PMTs' explanations about what can be done for this student in the case of being the teacher of him/her were depending on the interaction of teacher and student. For this case some of the explanations for other distracters were unavoidably considered as true intervention. So, this observation signified the necessity of distracter revision. Based on both interview and content analysis of responses, researcher decided to turn distracters of VP1 into complex structure of multiple-choice type (See Appendix B).

Table 4.1 Summary of Content Analyses of Open-Ended Responses

	Correct		*Incorrect*		
Item	*#*	*Mean*	*#*	*Mean*	*Change*
VP1	22	1.14	22	1.04	The structure of item (changed to CMI*)
LS2	19	1.53	25	0.64	Phrases and visual (Revised)
AP3	34	1.26	10	0.70	No change
AP4	5	1.46	39	1.15	The structure of item (changed to CMI*)
AP5	32	0.69	12	0.50	Phrases and visual (Revised)
LS6	8	1.38	36	0.14	Phrases (Revised)
LAP7	27	1.52	17	1.10	Distracters (Revised/combined with LP28)
MS8	13	1.10	31	0.74	Distracters (Revised)
VS9	27	0.61	17	0.24	Phrase (Emphasized)
AP10	7	1.43	37	0.65	Phrases and Distracters (Revised)
LS11	28	1.29	16	0.93	Distracters and visual (Revised)
LS12	28	0.82	16	0.56	Distracters and visuals (Revised)
AS13	18	0.56	26	0.38	Phrase (Emphasized)
LAP14	10	1.5	34	0.38	Phrase (Revised)
AP15	38	1.47	6	0.00	Visual (Revised)
LS16	14	1.5	30	0.24	No change
AP17	0	0.00	44	0.74	No change
VS18	29	1.72	15	0.67	No change
AS19	23	1.00	21	0.29	Phrases and visual (Revised)
MS20	17	1.12	27	0.52	Phrases and Distracters (Revised)
AP21	7	1.29	37	0.62	The structure of item (changed to SMI**)
AP22		Eliminated			
AS23	38	0.82	6	0.33	No change
VS24	15	0.93	29	0.10	Phrase (Revised)
LAS25	37	1.05	7	0.29	Visual (Revised)
AVP26	26	1.07	18	0.33	Distracters and visuals (Revised)
VS27	11	0.82	33	0.67	Distracters (Revised)
LP28	31	1.48	13	0.46	Distracters (Combined with LAP7)
LP29	20	1.35	14	0.58	Visual (Revised)

CMI*: Complex multiple-choice item format, SMI*: Simple multiple-choice item format

LS2 asked PMTs to compare the lengths of two security tapes on identical cylinders, one was vertically positioned, and the other was rolled around (See Appendix A).The main comment written in the open ended-field was based on the claim that the assumed π value should be given in the stem. According to them the result will be equal when the value of π is assumed as 3 otherwise one is longer than other one. Hence the result of the comparison between these two lines depends on the value of π. One PMT who stated the necessity of information related to the value of radius (r) of ball during the interviews explained the idea behind these comments. He claimed that without the r-value, this comparison is impossible. Based on this observation distracters worked well in terms of content, so there was no change made related to the content of distracters on LS2. Interview findings related to LS2 were explained in detail in Chapter III where the rubric for interview data was exemplified. To sum up, visual and wording of this item was revised (See Appendix B).

AP3 asked PMTs to diagnose the reason why student made mistake given in the stem, i.e. misconception of area, and asked them to determine the case, which might mislead students to make wrong generalizations on the use of area formula (See Appendix A). Item AP3 was one of the good working items and there was no change for this item.

AP4 asked PMTs to diagnose the reason why student made the mistake in the area comparison given in the stem and asked them to suggest an activity to help the student given in the case to overcome his difficulty (See Appendix A). Based on the open-ended responses and interview findings, the clear observation about this item was that majority of responses could be attributed to clear diagnosis of the student's difficulty but they failed to provide completely true suggestion. Although most of the participants were aware of the student's difficulty in measuring the area with non-uniform unit (rectangular unit), almost all of them suggested intervention of such a case with the uniform unit (1 square unit). However, as indicated in item stem student in that case had no difficulty with area measurement with uniform units. Not only open-ended responses but also interview findings indicated that some of PMTs' explanations about what can be done for this student in the case of being the teacher of him/her were depending on the interaction of teacher and student. For this case some of the explanations for other distracters were unavoidably considered as true intervention. Thus, this observation signified the necessity of distracter revision. So, based on this observation, researcher decided to turn distracters of AP4 into complex structure of multiple-choice type. Moreover in order to simplify the item stem, visuals and wording of the item were also revised (See Appendix B).

For the item AP5, there were 3 cases given to PMTs to decide whether they were appropriate for the evaluation of area concepts of rectangular prism, square prism and cube for grade 6 (See Appendix A). The first case was directly related to surface area and exists in 6^{th}grade elementary mathematics program. The second case was not directly related to calculating surface area. The third one was directly related to the surface area computation and generalizing the idea. However, the level of the item was not appropriate for 6^{th}grade level, instead it is given in the 8^{th}grade mathematics curriculum. As a matter of fact there is a warning in the elementary mathematics curriculum for teachers not to introduce this concept at 6^{th}grade level. However, almost all interviewees stated that they had no information about the warnings in the curriculum, and stated that they did not notice warnings when they examined curriculum in related courses. It was observed that PMTs had different approaches towards the content of the item, and to determine correct answer. Based on these interview findings, the content and statement of the Case II was changed. The content of the statement adapted to cube, which was more definite, and the phrase of "maximum" was eliminated. To sum up, in order to standardize the visuals of item all visuals of prerequisites of the item were eliminated, the statement of stem was changed to positive, statement of Case I was changed, and finally the statement and content of Case II was changed for further administrations (See Appendix B).

For the item LS6, there were 4 cases given to PMTs to call into question of four students' statements given on length of major and minor arc, which is introduced in 7^{th} grade Turkish elementary mathematics curriculum (See Appendix A).Although length of major and minor arc is introduced in 7^{th} grade Turkish elementary mathematics curriculum, almost all PMTs claimed that they had no idea about length of an arc on circle. Thus, they had no detailed explanation for this item. However, in order to simplify the stem there was small wording revisions for the second administration.

For the item LAP7, there were 3 cases given to PMTs to categorize approach (es) of students on length measurement cases similar to the wrong conceptualization of area measurement which was given in the stem of the item (See Appendix A). The first case was constructed similar to the student's approach given in the stem. In this example student assumed that the same number of units refers the same measurement quantity. In the second case, the student's approach was again related to the number of units and measurement quantity, but somehow, the underlying idea was quite different from the approach in the stem and also Case I. In other words, student's thinking on the same number of units does not imply to the bigger number of units refers to bigger measurement quantity as given in the Case II. The third case was about the measurement quantity and organization of units, which was quite different from all other cases. Responses on these alternatives had clear diagnosis of student's way of thinking given in the stem of the item, especially the number of units and the measurement quantity. But almost all responses stated that student in case I has also possibility to make error given in Case II. But the main point was that approaches of these two students were quite different and only 4 of PMTs emphasized this difference. So, this observation signified the necessity of distracter revision. As a result, in order to improve alternatives from this perspective, it was decided to combine these two items (LAP7 and LP28) since these two items focused on the similar measurement concepts (See Appendix B).

For item MS8, PMTs asked to determine the specific measurement, which was not directly related to any property of the bucket given in the stem (See Appendix A). There were 5 measurement cases given to PMTs to consider, and they were asked to specify the indirect measurement of any property of bucket. Item MS8 was one of controversial items in the test. Findings of interview data revealed that some further modifications needed for MS8 in addition to findings of open-ended responses. Interviewees suggested that there was a need to clarify some details of distracters. To begin with participants stated that the statement of Case I of MS8 was hard to understand. The second suggestion was about an uncertainty about the coefficient of elasticity for Case IV and properties of the coin in Case II. As a result, in order to improve alternatives from this perspective, it was decided to revise and change Statement of Case I, to insert elasticity coefficient in the statement of Case IV, and to insert the numerical quantity in the statement of Case II (See Appendix B).

For item VS9, PMTs asked to calculate the volume of the geometric object after revolved around x- axis (See Appendix A). The common response on this item revealed that almost all PMTs had difficulty on visualizing the geometric object after revolved with respect to the x-axis on coordinate system, and accordingly had no idea how to calculate the volume of the shape. On the other hand, any attempt to find the total volume of truncated cone and volume of cylinder without a cone inside was enough to reach a true result. The most striking observation related to this item was that there were PMTs who tried to calculate the volume by sliding some parts of 2-dimensional shape to make regular object. This approach completely ignored the conservation of volume and signaled the existence of misconception on conservation. The other observation was that there were two participants who confessed that they did not realize the information of revolution for 180^o with respect to x-axis. So it was necessary to underline this information in the stem. As a result, in order to improve VS9,it was decided to underline and to emphasize the information about the rotation angle of 180^0.

For item AP10, PMTs were asked to determine the prerequisites for calculation of the surface area of cone (See Appendix A). There were two interesting observations related to this item. Firstly, based on explanations of PMTs, it was observed that some of PMTs were confused about the volume and surface area concepts. Secondly, some of them failed to construct a net for cone and to formulize the surface area of cone. At this point some PMTs who tried to construct the net of cone considered surface area of cone as only sector of circle. Accordingly they thought that the concept of area of sector was sufficient for surface area of cone. There were only 5 PMTs who have clear conception of surface area of cone successfully and answered the item with clear reasoning. The possibility of PMTs to confuse the surface area and volume was not considered at item construction step. As a result for this item, content of distracters were revised and the distracter including I, IV was replaced one of existing ones -which confused surface area and volume- (See Appendix B).

LS11 asked PMTs to calculate the length of the path of an ant walking outside the cylinder (See Appendix A). The result can be calculated by Pythagoras formula easily, which was exactly the length between two points on the rectangular net of cylinder. The only comment on this item was on the weaknesses of visual of this item. Based on comments about the visual of this item, it was necessary to insert the exact location of ants and to make other information more definite for the further administration. Another comment was about the participants' difficulty in imagination about how ant walked through outside the cylinder. Based on this comment, walking route of the ant was given in the statement. Although majority of PMTs stated that they were used to such items before, there existed two PMTs having difficulty to find a strategy how to calculate the length. The common mistake done by majority of the PMTs was considering the path of ant as a total diagonal of net instead of half. The other interesting observation on this item was that there were some PMTs who attempted to use merely the radius of cylinder instead of half of the circumference. Based on this observation, one of the distracters was replaced with the result of the calculation with merely quantities of radius for further administration. Moreover, in order to improve alternatives from this perspective, it was decided to make visuals more clear. The routes on the cylinders were drawn in the item stem (See Appendix B).

LS12 required PMTs to compare the length of the curves inside circles (See Appendix A). This can be either calculated easily by assigning some values to radii or by different estimation strategies. The illustration of this item was commented on by most of the participants. Based on these comments the off-color points were revised, and the figure became clearer for further administrations. The common mistake done by majority PMTs was visual comparison of curves, which optical illusions of curves misled them. Finally for this item, some of PMTs stated that they did not need to calculate the third one given in the item. Since there was only one alternative which states I=II. This observation strongly signified to add new alternatives, which includes equality of other cases. Based on these observations, the visual of item was improved and alternatives were revised for the further administrations (See Appendix B).

For the Item AS13, 5 different strategies were given to PMTs to find the area of trapezoid and they were asked to decide on the generalizability of each strategy to all types of trapezoids (See Appendix A).Interview findings of this item revealed that PMTs have some defective approaches for generalizability of the given cases. The interesting observation was that most of the answers failed to come up with the conclusion that the strategy was true for the specific case but failed in generalizing for all trapezoids. This observation signified the necessity to emphasize the generalizability term in the stem of the item. The only suggestion for improvement of the AS13 was about the necessary information about the structure of trapezoids. There were few PMTs pointed at the necessity of statement on the characteristics of trapezoid, whether isosceles or not. Although it was not necessarily needed but in order to be more clear the phrase of isosceles trapezoid was added for further administrations. There was no other comment done for the structure of this item. As a result in order to improve AS13, the phrase of isosceles was inserted and the phrase of generalizability in the item stem was emphasized (See Appendix B).

Item LAP14 asked PMTs to diagnose the reason why student made mistake in the given case regarding the relationship between area and perimeter and asked them to determine the counter examples that could be provided for this case (See Appendix A).Responses on these alternatives had clear diagnosis of student's error given in the stem of the item. There was no suggestion, thus, there was no change for this item.

Item AP15 requested PMTs to diagnose the student error in the given case in relation to the rectangular areas, and to estimate the answer of student if a similar case is given (See Appendix A). Item AP15 was one of the easy items in the test. Almost all PMTs answered this item easily. There was clear information for diagnosing the problem stated in stem or complete explanation for the answer. The main observation related to this item was that interviewees' conception of easiness of this item comparing to other items in the test. Almost all PMTs stated that Item AP15 was one of the easiest items in the test. Some of them attributed the easy structure of this item to the easy realization of the reason why student made mistake in the given case and to the computational structure of the item. According to them, they feel comfortable when they make computations for the solutions. There were only two interviewees who failed to diagnose the student's thinking and failed to determine how he reached such a result. As a result, there was no need to change, but in order to shorten one of the visuals was deleted (See Appendix B).

Item LS16 asked PMTs to estimate the length of the line rolled around a cylinder200 times (radius of cylinder was given). The result can be estimated with different strategies (See Appendix A). The first one can be finding the maximum value and minimum value then calculating the average value. The second one can be performing calculations based on the average radius value. The main observation interview finding on this item was that interviewees' conception of difficulty level of this item. Almost all interviewees stated that LS16 was one of the most difficult items in the test. Except the statements on difficulty level of this item, there was no suggestion related to this item. Participants had clear understanding on what was asked for this item. Thus, there was no need to make any change related to content of either stem or distracters of LS16.

Item AP17 asked PMTs to verify the proof of Pythagoras Theorem by using area of different geometric shapes (See Appendix A).There were no suggestion or revision related to this item. Thus, there was no need to make any change related to content of either stem or distracters of AP17.

Item VS18 requested PMTs to determine the volume of an irregular object (See Appendix A). There were no suggestion or revision related to this item. Thus, there was no need to make any change related to content of either stem or distracters of VS18.

For item AS19, three pairs of area quantities given to compare in the item stem and PMTs were asked to elaborate the use of standard units in area measurement (See Appendix A).Majority of participants explained that they believed that all unit squares had the same quantity and they all should be equal. This observation signified the necessity of some phrases such as “units are at varying size and shape” in the item stem. For this reason the statement of the item and the description of the case given in the item stem was revised. Moreover in order to simplify the item, item phrase was revised and visual was eliminated (See Appendix B).

The item MS20 requested PMTs to determine the estimation made by a different strategy and consisted of the concepts of area, weight and volume (See Appendix A). There was no suggestion for this item. However, distracter phrases were revised and clarified.

For the itemAP21, 4 cases were given to PMTs to decide whether they were appropriate for the evaluation of area concepts for 6^{th}grade level (See Appendix A). The first case was directly related to conversion of area units. The second case was about area measurement by using unit squares. The third one required calculating the area of circle, which is introduced at 7^{th}grade in elementary mathematics curriculum. Moreover there is a warning for teachers not to introduce this concept at 6^{th}grade level in the curriculum. The last one was about the estimation of irregular area on map. There was only one explanation suggesting reorganization of distracters of this item. The explanation pointed out that there was only one alternative, which excluded case III. Since the participant knew that area of circle is 7^{th} grade concept in curriculum, she was quite sure about her answer. This information signified the necessity to revise distracters. However, considering findings the structure of this item was turned into simple multiple-choice item, and prerequisites of the item were revised as distracters (See Appendix B).

Item AP22 demanded PMTs to diagnose the reason why student made mistake in the calculation of area as mentioned in the stem and asked them to suggest an activity to help such student overcome this difficulty (See Appendix A). Responses on these alternatives had clear diagnosis of student’s error given in the stem of the item. But some of PMTs’ explanations about what can be done for this student in the case of being the teacher of him/her were depending on the interaction of teacher and student. For this case some of the explanations for other distracters were unavoidably considered as true intervention. Thus, this observation signified the necessity of both stem and distracter revision. However this item was eliminated for the second administration.

The item AS23 asked PMTs to verify the area calculation strategy and check the generalizability of this strategy (See Appendix A). According to results, item AS23 was one of the easy items in the test. Almost all PMTs answered this item easily. Most of participants explained how they think for this item clearly. They provided a clear mathematical explanation for their reasoning. As a result, based on these observations distracters worked well in terms of content, so nothing was changed related to AS23.

The item VS24 asked PMTs to calculate the number of balls that can be put inside a given package (See Appendix A).In other words, they were asked to find the volume of a given package in terms of ball, as a unit. Since it was an estimation problem, it was given approximate values for the circumference of the balls instead of just one value. The result of this item could be either calculated easily by using the dimensions given in the stem, or by other estimation strategies. For the initial version of VS24, approximate values of ball circumference made participants to feel uncomfortable about the result. Thus one of the values was eliminated, since the result did not change with the approximate values of circumference. The other observation related to this item was that about one third of participants stated that they had difficulty to understand the meaning of circumference of a ball. According to them circumference is directly related to two-dimensional shapes and three-dimensional objects may not have circumference. Based on this observation, for the further administration the “circumference of ball” was defined as “the circumference of the largest circle in the ball” (See Appendix B).

The Item LAS25 asked PMTs to compare the total circumferences and total areas of two circles and three circles, which were made up with the same length of rope (See Appendix A). The aim of the item was only to compare total circumferences of first two with the total circumferences of the next three. The expected answer was only the total circumferences of first two is equal to total of three but the total area of the first two is more than total area of next three, which is pretty obvious. But, some pre-service mathematics teachers stated in the interview that they tried to compare areas and circumferences of these two and three circles based on visual of the LAS25. In other words, they tried to make some estimation about overlapping areas, which made the item much more complicated than the situation, which no researcher or experts had considered before. This observation required revising the visual of item as separate sets of circles (See Appendix B).

For the Item AVP26, there were 4 cases regarding deformation of the area of circle based on the formula of area of parallelogram given to PMTs to diagnose student's difficulty and determine similar cases, which had possibility for students to struggle in the same way (See Appendix A). The first case was direct deformation of object and had no idea of convergence. The second case was about constructing a sphere with cones of equal size, which had idea of convergence. The third one was again focused on deformation of area. Final one was related to direct measurement and had no idea of convergence. Responses on this item had clear diagnosis of student's difficulty on the idea of convergence. But almost all of them failed to catch the idea that the third one was related to deformation of area instead of the idea of convergence. Another observation on the test booklets was that many PMTs eliminated the distracters including the Case IV. This might imply that they easily came up with the idea of direct measurement given in Case IV. Based on this observation and also in order to shorten the item, Case IV was eliminated for next administrations. Furthermore, in order to be more definite visuals were inserted for each case (See Appendix B).

For the item VS27, there were 4 cases given to PMTs to diagnose student difficulty on volume measurement (See Appendix A). This item was one of the controversial items in the test. The interesting observation on this item was all PMTs tried to diagnose the student's mistake; however they failed to provide completely true suggestion. There was only one response, which provided completely true explanation for this case. He had clear diagnosis of student's mistake on the volume measurement, and he determined the reason why student was mistaken. The rest of responses were either null or completely wrong comment, or lack of information for diagnosing the stated problem. During the interviews PMTs stated different approaches why the case in the item was wrong. Based on the interview findings major approaches were categorized into five and were replaced with distracters of the item (See Appendix B).

For the item LP28, there were 4 cases given to PMTs to categorize students' way of thinking on length measurement (See Appendix A). Even though responses on these alternatives included clear diagnosis of students' thinking approach, there exist two parallel items in the test. Thus Item LAP7 and Item LP28 were combined for the second administration (See Appendix B).

For the item LP29, there were 4 cases given to PMTs to diagnose students' mistake when calculating the perimeter of a trapezoid (See Appendix A). All complete and mathematically true reasoning are clustered on the correct answer of multiple-choice item (E). All responses pointed out the correct answer as (E) explained the reason why student was mistaken while calculating the perimeter of the shape. But the rest of responses with 1 point score which could be attributed to lack of information or incomplete explanation for diagnosing students' way of thinking. Most of them failed to determine the reason why student made mistake. As a result, distracters worked well in terms of content, so there was no change made related to the content of distracters on Item LP29. In order to make the item more clear, the type of the papers used as background in the distracters were inserted for the next administrations (See Appendix B).

As a result, as seen in the Table 4.1, there were 5 items with no change, 11 items had minor revisions either phrases or visuals of the item. Distracters of 7 items were revised; in some cases they were re-written. The structure of 3 items, were changed; two items were re-constructed in a complex multiple choice item format, and one item vice-versa. Finally 1 item was decided to eliminate for further administrations.

4.2 Implementation of the Test and Results of Field Testing II

The purpose of the second administration was to understand how items worked from the perspective of both Item Response Theory and Classical Test Theory. So both

Item analysis and Rasch analysis were conducted. In the first round of the field-testing, there were 29 multiple-choice items. Based on findings of Round 1, there were remaining 26 items. It was decided to split retaining items into two parallel tests in order to shorten the administration time. In the Round 1 it was observed that that the administration time was long for one lesson hour. Thus, in the Round 2 there were two tests having 16 items with 6 of them were in common (AP4, MS8, LAP1, MS20, AVP26, and VS27). Both Test 1 and the Test 2 were covering the concepts based on mathematics knowledge for teaching measurement.

For each item, item difficulty, and fit statistics were calculated with the help of the Rasch Model. Moreover, the difficulty, discrimination index, correlation with the total scores, and point biserial correlation values were determined from the Classical Test Theory perspective.

4.2.1 Results of Rasch Analyses of Field Testing II

The binomial Rasch model was used to a used to examine the extent to which the test measured one attribute at a time on a hierarchical line of inquiry. Once the data set had been created, the Rasch Analysis was conducted using BIGSTEPS computer software (Wright & Linacre, 1991).

4.2.1.1 Unidimensionality of Test 1 and Test 2

Through use of fit statistics generated by Rasch analysis, unidimensionality of Test 1 and Test 2 were assessed. Rasch Analysis provides two types of mean square fit statistics: infit and outfit. Infit refers to fit statistic which is more sensitive to unexpected patterns of observations by persons on items that are roughly estimated. On the other hand, outfit refers to outlier-sensitive fit statistic and it is more sensitive to unexpected observations by persons on items that are relatively very easy or very hard for them (Linacre, 2007).Acceptable values for mean squares range from 0.7 to 1.3 (Linacre, 2007). Figure 4.1displays the item statistics of Test 1. All of the items had infit/outfit statistics within the acceptable range of 0.7 and 1.3, indicating that the test measured a unidimensional construct

ITEMS STATISTICS: MISFIT ORDER

ENTRY NUMBR	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBIS CORR.	ITEMS
9	135	473	-.06	.10	1.05	1.1	1.06	.9	A .01	AP5
12	67	470	.84	.13	1.06	.7	1.05	.3	B-.02	VS27
15	209	430	-.96	.10	1.03	1.0	1.06	1.4	C .06	LS12
11	141	473	-.13	.10	1.05	1.1	1.06	.9	D .02	MS8
13	79	465	.63	.12	1.02	.2	1.04	.3	E .05	VS18
2	268	473	-1.32	.09	1.03	.8	1.03	.6	F .05	AP3
10	149	472	-.21	.10	.99	-.2	1.02	.3	G .09	AVP26
16	149	373	-.56	.11	.98	-.4	1.00	.1	H .12	VS24
5	113	473	.19	.11	1.00	-.1	.98	-.2	h .09	AS19
8	18	473	2.25	.23	.99	.0	.89	-.4	g .07	AP17
6	91	472	.47	.11	.99	-.1	.98	-.2	f .09	LAP14
7	203	473	-.73	.09	.98	-.7	.96	-1.0	e .12	MS20
4	42	473	1.37	.16	.97	-.3	.77	-1.4	d .15	AP4
3	400	473	-2.82	.12	.96	-.5	.87	-1.1	c .12	AP15
14	59	455	.96	.14	.96	-.4	.81	-1.4	b .16	AP10
1	122	473	.08	.10	.95	-1.0	.94	-.8	a .17	LS2
MEAN	140.	462.	.00	.12	1.00	.1	.97	-.1		
S.D.	92.	25.	1.14	.03	.03	.6	.09	.8		

Figure 4.1 Item Measure Information of Test 1

ITEMS FIT GRAPH: MISFIT ORDER

```
+---------------------------------------------------------------------------+
|ENTRY| MEASURE | INFIT MEAN-SQUARE   | OUTFIT MEAN-SQUARE  |        |
|NUMBR| -     + |0     0.7 1 1.3     2|0     0.7 1 1.3     2| ITEMS  |
|-----+---------+---------------------+---------------------+--------|
|   9 |     *   |      :   *   :      |A     :   *   :      | AP5    |
|  12 |    *    |      :   *   :      |B     :   *   :      | VS27   |
|  15 |     *   |      :  *.   :      |C     :   *   :      | LS12   |
|  11 |  *      |      :   *   :      |D     :   *   :      | MS8    |
|  13 |  *      |      :   *   :      |E     :   *   :      | VS18   |
|   2 |   *     |      :   *   :      |F     :   *   :      | AP3    |
|  10 |       * |      :   *   :      |G     :   *   :      | AVP26  |
|  16 |   *     |      :  *.   :      |H     :   *   :      | VS24   |
|   5 |  *      |      :   *   :      |h     :  *.   :      | AS19   |
|   8 |    *    |      :   *   :      |g     :  *.   :      | AP17   |
|   6 |      *  |      :  *.   :      |f     :*  .   :      | LAP14  |
|   7 |     *   |      :  *.   :      |e     :  *.   :      | MS20   |
|   4 |    *    |      :  *.   :      |d     :  *.   :      | AP4    |
|   3 |    *    |      :  *.   :      |c     :  *.   :      | AP15   |
|  14 |    *    |      :  *.   :      |b     :  *.   :      | AP10   |
|   1 |*        |      :  *.   :      |a     :*  .   :      | LS2    |
+---------------------------------------------------------------------------+
```

Figure 4.2 Item Fit Information for Test 1

. In addition, item difficulties ranged from -2.82 (AP15) to 2.25 (AP17). Figure 4.2 also provides a visual display of findings.

Similarly, Figure 4.3displayed the measure information about items in Test 2 and Figure 4.4 displayed the item fit information about in Test 2 respectively. As seen in the figures, similar to Test 1, all items in Test 2 were between the critical values of .7 and 1.3, thus there were no misfit items. Difficulties of items of Test 2 ranged from (-4.20 (AP13) to 1.06 (MS20).

ITEMS STATISTICS: MISFIT ORDER

ENTRY NUMBR	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBIS CORR.	ITEMS
14	379	395	-4.20	.25	1.13	.5	.93	-.2	A .03	LAS25
9	119	383	-.07	.11	1.05	1.0	1.05	.6	B-.07	AVP26
16	127	389	-.15	.11	1.03	.7	1.01	.1	C-.05	AP4
3	80	398	.51	.12	1.03	.4	1.02	.1	D-.04	AP13
11	150	395	-.38	.10	1.02	.5	1.01	.2	E-.02	LAP7
7	95	395	.28	.11	1.01	.2	.98	-.2	F-.02	AP21
10	50	396	1.06	.15	1.00	.0	.90	-.6	G-.01	VS27
8	172	396	-.63	.10	1.00	.2	.97	-.5	H .00	MS20
15	88	356	.23	.12	1.00	-.1	.96	-.4	h .01	LS16
5	64	396	.78	.13	.99	-.1	.94	-.4	g .02	LAP14
6	37	389	1.37	.16	.99	-.1	.83	-.9	f .05	LS11
1	53	394	.99	.14	.98	-.1	.91	-.6	e .03	VP1
4	67	301	.37	.13	.98	-.2	.95	-.4	d .04	VS9
13	122	393	-.07	.11	.98	-.4	.98	-.3	c .02	MS8
12	133	393	-.21	.10	.97	-.6	.94	-.9	b .04	LP29
2	99	363	.12	.11	.97	-.6	.94	-.7	a .05	LS6
MEAN	115.	383.	.00	.13	1.01	.1	.96	-.3		
S.D.	78.	24.	1.21	.04	.04	.4	.05	.4		

Figure 4.3 Item Measure Information of Test 2

```
                  ITEMS FIT GRAPH:  MISFIT ORDER

+----------------------------------------------------------------------+
|ENTRY| MEASURE  |  INFIT MEAN-SQUARE  |  OUTFIT MEAN-SQUARE  |        |
|NUMBR| -      + |0      0.7 1 1.3    2|0      0.7 1 1.3     2| ITEMS  |
|-----+----------+---------------------+----------------------+--------|
|   14|*         |        :  .* :      |A      : *.   :       | LAS25  |
|    9|      *   |        :  *  :      |B      :  *   :       | AVP26  |
|   16|      *   |        :  *  :      |C      :  *   :       | AP4    |
|    3|       *  |        :  *  :      |D      :  *   :       | AP13   |
|   11|     *    |        :  *  :      |E      :  *   :       | LAP7   |
|    7|      *   |        :  *  :      |F      : *.   :       | AP21   |
|   10|        * |        :  *  :      |G      :*  .  :       | VS27   |
|    8|     *    |        :  *  :      |H      : *.   :       | MS20   |
|   15|      *   |        : *.  :      |h      : *.   :       | LS16   |
|    5|       *  |        : *.  :      |g      : *.   :       | LAP14  |
|    6|        * |        : *.  :      |f      :*  .  :       | LS11   |
|    1|       *  |        : *.  :      |e      : *.   :       | VP1    |
|    4|       *  |        : *.  :      |d      : *.   :       | VS9    |
|   13|      *   |        : *.  :      |c      : *.   :       | MS8    |
|   12|      *   |        : *.  :      |b      : *.   :       | LP29   |
|    2|      *   |        : *.  :      |a      : *.   :       | LS6    |
+----------------------------------------------------------------------+
```

Figure 4.4 Item Fit Information for Test 2

4.2.1.2 Item Difficulty of Test 1 and Test 2

The initial step was to analyze the extent to which the results were consistent with the researcher's intentions, as emphasized by Wright and Stone (1979). In order to ensure the accuracy of ranking in the Rasch analysis, observed and estimated item difficulty statistics were correlated. The observed rankings were then compared to the theorized rankings using Spearman's Rank Order Correlation.

The first analysis using all items from the data of Round 2, the Spearman Rank Correlations was r= .989 (for Test 1) and r= .982 (for Test 2). The correlation value suggested that the ordering of the observed item mostly confirmed the theoretical expectation. Items that were expected to be easily endorsed by respondents had item locations indicative of an easy item endorse, while more difficult items yielded locations indicative of a difficult item endorse.

4.2.1.3 Item Person Map of Test 1 and Test 2

In order to determine the order of item difficulty in relation to the distribution of person ability and to identify item gaps, the person-item map was visually inspected. Figure 4.5 presents the distribution map of the items and persons. All items were evenly distributed over -3 and +3 range. Since the items were not clustered at one point, the chance factor of the answers and the possibility of the problem related to the key had been eliminated. On the other hand, the distribution of person was between (-4, +1), indicating that person measures varied between (-4, +1). Range of item difficulties and person abilities indicated that the items locations were higher than the person locations. That is, item difficulties are beyond the student abilities.

Among the items, AP17lined at the top, whereas AP15lined at the bottom. That means, AP17 was the most difficult item and AP15 was the easiest item in the Test 1. Except AP17 and AP4, the distribution of the item locations extended along the continuum of person locations. This spread of item locations is desirable as it allows for differentiation among participants through the use of item difficulties. As AP17 was found to be too difficult, decision was to delete that item for further administration.

```
           MAP OF PERSONS AND ITEMS
                <frequ> |  <less>
  3                     +
                        |
                        |
                        |
                        |
                        |
                        |Q AP17
                        |
  2                     +
                        |
                        |
                        |
                        |
                        |  AP4
                        |
                        |S
  1                   . +  AP10
                        |  VS27
                        |
                     .# |  VS18
                    . Q|   LAP14
                     .# |
                      . |  AS19
                      . |  LS2
  0                 .## +M AP5
                      . |  MS8
               .####### |  AVP26
                    . S|
                      . |
                .###### |  VS24
                     .# |  MS20
                      . |
 -1        .########### +  LS12
                       M|S
                     .# |
           .########### |  AP3
                     .# |
                        |
                    .## |
               .####### S|
 -2                     +
                      . |
                        |Q
                  .#### |
                        |
                       Q|
                        |
                        |  AP15
 -3                     +
                      . |
                        |
                      # |
                        |
                        |
                        |
                        |
 -4                   . +
                  <rare>|<more>

EACH '#' IN THE PERSON COLUMN IS   3 PERSONS; EACH '.' IS 1 TO   2 PERSONS
```

Figure 4.5 The distribution map of items and persons of Test 1

The mean of the test scores was 4.69 (*SD*=1.99) out of 16 and median was 5.0. Low mean score indicated that test items were difficult for participants. As seen in Figure 4.5, the scores were normally distributed (Skewness = .402, and Kurtosis =0.19).

Similar to Distribution map of Test 1, items of Test 2 evenly distributed over (-3,+2) range as seen in the Figure 4.6 and items were not clustered at one point. On the other hand, the distribution of person was between the range of (-4, +1). Overall the results of the variable maps indicated that the items locations were higher than the person locations. At this point, the range of person abilities was below the range of item difficulties, which means that Test 2 was also difficult for the examinees. LS11 was at the top, whereas LAS25 was at the bottom of the vertical line. LS11 was the most difficult, and LAS25 was the easiest item in Test 2. The rest of the items evenly distributed along the vertical line. Moreover, two pairs of items (LAP14 and VP1) and (MS8 and AP4) in the Test 2 had same item measures.

```
MAP OF PERSONS AND ITEMS
             <more> | <rare>
 2                  +
                    |
                    |T
                    |
                    |
                    |
                    |
                    |  LS11
                    |
                    |
 1                  +
                    |S VS27
                    |  VP1  LAP14
                    |
                 .  |
                 .  |  AP13
                 . T|
                    |  VS9
                 .  |  AP21
                .#  |  LS16
 0               .  +M LS6
                 .  |  AVP26
              .###  |  MS8  AP4
                 #  |  LP29
               .## S|
        .#########  |  LAP7
                .#  |
                .#  |  MS20
      .###########  |
                 #  |S
-1              .#  +
              #### M|
      .###########  |
                 .  |
                 .  |
                .#  |
              .###  |
          .#######  |
                 . S|T
                 .  |
-2               .  +
                 .  |
                 .  |
             .####  |
                    |
                   T|
                    |
                    |
                    |
                    |  LAS25
-3                  +
                 .  |
                ##  |
                    |
                    |
                    |
                    |
                    |
-4               .  +
            <rare> | <more>
EACH '#' IN THE PERSON COLUMN IS   6 PERSONS; EACH '.' IS 1 TO   5 PERSONS
```

Figure 4.6 The distribution map of items and persons of Test 2

The score distribution of participants on Test 2 was displayed in Figure 4.6. The mean of the test scores was 4.31 (*SD*=1.72) out of 16 and median was 4.0. Low mean score indicated that test items were difficult for participants. Similar to Test 1, low mean score indicated that test items of Test 2 were difficult for participants. As seen in the Figure 4.6 the scores were normally distributed (Skewness = .119, and Kurtosis = .109).

4.2.1.4 Reliability and Separation Indices of Test 1 and Test 2

Rasch analysis generates two reliability indices: one for item separation and one for person separation. A value of 1 represents high separation, in which errors are low and item difficulties and students' measures are well separated along the scale (Wright & Masters, 1981). Figure 4.7 provided item summary statistics. In Rasch analysis, the extreme scores are excluded. Therefore, Rasch analysis was conducted with 473 non-extreme persons.

SUMMARY OF 16 MEASURED ITEMS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	140.3	462.1	.00	.12	1.00	.1	.97	-.1
S.D.	92.3	25.4	1.14	.03	.03	.6	.09	.8
MAX.	400.0	473.0	2.25	.23	1.06	1.1	1.06	1.4
MIN.	18.0	373.0	-2.82	.09	.95	-1.0	.77	-1.4

REAL RMSE	.12	ADJ.SD	1.13	SEPARATION	9.12	ITEM RELIABILITY	.99
MODEL RMSE	.12	ADJ.SD	1.13	SEPARATION	9.17	ITEM RELIABILITY	.99
		S.E. OF ITEM MEAN	.29				

Figure 4.7Summary of Item Information of Test 1

Findings indicated that there was a good item separation as item reliability is higher than .90. Moreover, items in the Test 1 were separated into almost nine difficulty levels (with a Separation of 9.12).The mean and standard deviation of the Infit and Outfit for items indicated that these values were very close to expected values. As seen in Figure 4.7, mean square value of Infit is 1.00 while mean square value of Outfit is .97. Overall, these measures indicated that the items in the Test 1 fit the model well.

Moreover, Rasch Analysis conducted with 506 non-extreme persons for Test 2. The following figure summarized the item separationindices.

SUMMARY OF 16 MEASURED ITEMS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	136.3	482.6	.00	.11	1.00	-.1	1.01	-.1
S.D.	76.7	30.0	.76	.01	.02	.5	.08	.9
MAX.	392.0	501.0	.96	.13	1.04	.7	1.14	1.6
MIN.	60.0	380.0	-2.41	.09	.96	-.9	.88	-1.5

REAL RMSE .11	ADJ.SD .75	SEPARATION 6.97	ITEM RELIABILITY .98			
MODEL RMSE .11	ADJ.SD .75	SEPARATION 7.01	ITEM RELIABILITY .98			
S.E. OF ITEM MEAN = .20						

UMEAN=.000 USCALE=1.000

Figure 4.8 Summary of Item and Person Information of Test 2

Figure 4.8 summarized item information of Test 2. As seen in the Figure 4.8, similar to Test 1, there was good item separation (r=.98 ≥ .90) in Test 2. Items in the Test 2 were separated into almost six difficulty levels (with a Separation of 6.97).The mean and standard deviation of the Infit and Outfit for items indicated that these values were very close to expected values. As seen in Figure 4.8, mean square value of Infit is 1.00 while mean square value of Outfit is .88. Overall, these measures indicated that the items in the Test 2 also fit the model well.

4.2.2 Results of Classical Test Theory Analyses - Test 1& Test 2

Item analyses serve to identify statistically effective test items. For each item, the difficulty, discrimination index, and correlation values with the total score were determined.

4.2.2.1 Item Statistics of Test 1

Although data were collected from 504 participants, two outlier participants were eliminated. Thus, analysis was conducted based on test scores of 502 participants on Test 1, which consisted 16 multiple-choice items.

In order to control for guessing factor, participants were asked to answer how much confident they are about their answers on a five point rating scale for each item ("1"=Very unconfident, "2"=Unconfident, "3"=Neutral, "4"=Confident, and "5"=Very Confident) (See Appendix C). Majority of participants evaluate the correctness of their answers above the neutral value (*M*= 3.64, Min = 2.56, Max = 4.10) for each item. Table 4.2 presents item difficulty indices, two types of item discrimination indices, and percentage of examinees selecting each alternative. Each item was reviewed following the item analysis. Table 4.2 summarizes the item statistics for determining the flawed items and alternative statistics for diagnosing the examinee responses. Items flagged for statistical reasons to identify possible problems were identified. The review process also provided feedback that helped decrease the occurrence of poor quality items in further administrations.

Item Difficulty

Considering item difficulty (p) values in ITEMAN results, indices of all of items were quite low ranging from (p= .05) to (p=.76). Mean difficulty is .30, which can be interpreted that Test 1 was hard for examinees. Besides, only three items of which difficulty levels were considered as ideal, which is about or slightly above .5. These items were: Items AP3 (p= .55), MS20 (p= .44), and LS12 (p= .51). Haladyna (2004) points out that when the item difficulty was lower than .40, that item is considered as hard. Based on this criterion, 12 items were considered as difficult items, ranging from .05 to .36.

Item Discrimination Indices

Two indices were considered to assess item discrimination: D values and point biserial correlation. D values of .40 and greater are deemed as very good items, while below .19 are considered poor items (Ebel & Frisbie, 1986). According to Crocker and Algina (1986), the minimum acceptable point biserial value for this sample of 502 is 0.09. All of the items were found to have point biserial indices above .09. On the other hand, six items had an acceptable point biserial value (discrimination values between .09

and .3) but low discrimination index D. These items (AP4, LAP14, AP17, VS27, VS18 and AP10) along with the items previously flagged for unsatisfactorily low p-values, were reviewed individually to determine whether they should be kept, revised, or eliminated for future administrations of the instrument. For this case point biserial values, those were based on examinee responses, were preferred to be used for item selection.

Table 4.2 Item Analysis Results from 502 Examinees on 16 Item Test 1 (T1)

Item Statistics				Alternative Statistics (%)					
Item	*p*	*D*	*r*	A	B	C	D	E	Omit
LS2	0.25	0.40	0.52	0.25+	0.18	0.18	0.35	0.03	0.01
AP3	0.55	0.38	0.40	0.27	0.03	0.02	0.55+	0.13	0.00
AP15	0.76	0.41	0.55	0.02	0.02	0.06	0.04	0.76+	0.10
AP4	0.10	0.16	0.41	0.13	0.30	0.10+	0.02	0.45	0.00
AS19	0.23	0.32	0.45	0.15	0.49	0.03	0.09	0.23+	0.01
LAP14	0.19	0.23	0.37	0.16	0.18	0.22	0.19+	0.24	0.01
MS20	0.44	0.41	0.41	0.44+	0.09	0.09	0.17	0.20	0.01
AP17	0.05	0.07	0.34	0.28	0.08	0.37	0.05+	0.20	0.01
AP5	0.28	0.31	0.34	0.28+	0.07	0.02	0.11	0.52	0.00
AVP26	0.27	0.35	0.48	0.33	0.27+	0.22	0.10	0.06	0.02
MS8	0.33	0.32	0.37	0.04	0.39	0.17	0.33+	0.06	0.00
VS27	0.12	0.11	0.24	0.40	0.29	0.12+	0.08	0.10	0.01
VS18	0.15	0.24	0.48	0.22	0.10	0.36	0.15+	0.15	0.02
AP10	0.14	0.22	0.39	0.47	0.14+	0.13	0.21	0.04	0.00
LS12	0.51	0.39	0.38	0.18	0.14	0.04	0.09	0.51+	0.04
VS24	0.36	0.41	0.46	0.14	0.14	0.20	0.36+	0.05	0.11

+: The keyed response

Summary

The result of item analysis indicated that the alternative statistics of items indicated that all distracters of all items were good functioning (See Table 4.2). Despite their high difficulty, items LS2 (D= .40, *r*= .52), VS24 (D= .41, *r*= .46), were functioning well in terms of their discrimination indices. On the other hand, AS19 (D= .32, *r*= .45), AP5 (D= .31, *r*= .34), AVP26 (D= .35, *r*= .48), and MS8 (D= .32, *r*= .37), were considered as reasonable and needed minor revisions. There was only one item that could be assumed as easy AP15 (*p*= .76), and discrimination index (D= .41) of this item was good enough and all distracters of this item were functioning.

Based on this point biserial criteria for the discrimination there were 10 items (LS2, AP3, AP15, AS19, AS20, AP5, AVP26, LS12, and VS24) that were higher than both discrimination values. Such items would be in the test without any improvement. Again based on the Ebel's criteria items AP3, AS19, AP5, AVP26, MS8, and LS12 were reasonably good functioning but possibly needed minor revision to become well.

Finally, there were too difficult items in the test to be considered critically; AP4 (*p*= .10), LAP14 (*p*= .19), AP17 (*p*= .05), VS27 (*p*= .12), VS18 (*p*= .15), and AP10 (*p*= .14). At this point, items AP4, AP10 were decided to retain in the test since these items were targeting pedagogical content knowledge. Items AP17 and VS27 were poor items in terms of item discrimination. These were the most difficult items in the test, as expected and they had low discrimination indices. However, VS27 covered fundamental idea volume measurement. For this reason, this item was decided to retain in the test. But AP17 and LAP14 were decided to eliminate for further implementations.

To sum up, based on the item analysis results, there were five (LS2, AP3, AP15, LS12, VS24) items that could be retained in test as they were 4 items (AS19, AP5, AVP26, MS8) needed minor revisions, and one item (AP10) should be revised marginally, and three items (LAP14, MS20, and AP17) were eliminated for the Test 1. One of them AP17 (*p*= .05, D= .07) was eliminated considering item difficulty and other discrimination indices. Despite their acceptable values items, LAP14 and MS20 were also eliminated in order to decrease the number of items without disturbing the distribution of the content. Since there were other good-functioning items parallel to items; LAP14 and MS20.

4.2.2.2 Item Statistics of Test 2

Similar to Test 1, each item in Test 2 was reviewed following the item analysis. Table 4.3 summarizes the item statistics for determining the flawed items and alternative statistics for diagnosing the examinee responses. Item analysis was conducted based on 506 participants for Test 2 containing 16 multiple-choice items. Descriptive analysis indicated that majority of participants were confident about correctness of their answers (*M*= 3.50, Min=2.45, Max=4.01) Similar to Test 1, low mean score indicated that test items of Test 2 were difficult for participants.

Table 4.3 summarized item difficulty indices, two types of item discrimination indices, and percentage of examinees selecting each alternative. Each item was reviewed following the item analysis.

Item Discrimination Indices

According to Crocker and Algina (1986), the minimum acceptable point biserial value for this sample of 506 is 0.088. All of the items were found to have point biserial indices above .088. Therefore, items with point biserial coefficients greater than .088 may be retained if other predetermined item criteria are satisfactorily met. Although all items were found to have point biserial indices above .088, four items had discrimination values between .088 and .3.These items (VP1, AS13, LS11, and VS27) were reviewed individually to determine whether they should be kept, revised, or eliminated for future administrations of the instrument. On the other hand, there were five items (LS6, AP21, MS8, LAS25, and LS16), which has an acceptable point biserial value but low discrimination index D. For this case point biserial values were preferred to use for item selection. So, items AP21, MS8, LAS25, and LS16were decided to retain in the test. But, the item LS6 was directly related to the specialized content knowledge and the number of items related to this knowledge domain was quite enough. Thus, LS6 was eliminated for the next administration. Similarly, items VS9, LAP14, and LS11 were eliminated just because of low discrimination indices. As stated before MS20 was one of the common items in both Test 1 and Test 2. As mentioned in item analysis results of

Test 1, this item was directly related to specialized content knowledge. Although, the item had acceptable difficulty and discrimination index, this item previously decided to eliminate for the next administrations. Since there was other better functioning item parallel to MS20.Considering Ebel's criteria, items LAP7, LP29, and AP4 were reasonably good functioning but possibly needed minor revision to become well.

Table 4.3 Item Analysis Results from 506 Examinees on 16 Item Test 2 (T2)

Item Statistics				Alternative Statistics (%)					
Item	*P*	*D*	*r*	A	B	C	D	E	Omit
VP1	0.15	0.14	0.25	0.30	0.32	0.08	0.13	0.15+	0.01
LS6	0.24	0.27	0.45	0.30	0.21	0.13	0.04	0.24+	0.09
AS13	0.19	0.14	0.26	0.42	0.02	0.15	0.19+	0.21	0.01
VS9	0.17	0.18	0.38	0.20	0.17+	0.12	0.18	0.09	0.25
LAP14	0.15	0.19	0.41	0.16	0.25	0.19	0.15+	0.25	0.01
LS11	0.12	0.11	0.26	0.35	0.06	0.29	0.12+	0.15	0.03
AP21	0.23	0.16	0.30	0.08	0.44	0.23+	0.03	0.20	0.02
MS20	0.40	0.36	0.43	0.40+	0.09	0.07	0.17	0.27	0.01
AVP26	0.28	0.18	0.32	0.25	0.28+	0.27	0.09	0.08	0.04
VS27	0.16	0.11	0.20	0.39	0.25	0.16+	0.08	0.11	0.01
LAP7	0.35	0.31	0.41	0.35+	0.15	0.25	0.18	0.06	0.02
LP29	0.32	0.33	0.45	0.14	0.11	0.27	0.16	0.32+	0.02
MS8	0.29	0.22	0.33	0.06	0.40	0.14	0.29+	0.09	0.01
LAS25	0.77	0.27	0.44	0.08	0.03	0.04	0.77+	0.07	0.02
LS16	0.22	0.24	0.37	0.15	0.14	0.22+	0.31	0.07	0.11
AP4	0.29	0.30	0.40	0.13	0.30	0.29+	0.05	0.21	0.03

+: The keyed response

Item Difficulty

Considering item difficulty (p) values in ITEMAN results, indices of all of items were quite low ranging from (p= .12) to (p=.77) ; Mean Difficulty=.27, which can be interpreted as Test 2 was hard for examinees. Haladyna (2004) points out that when the item difficulty was lower than .40, that item is considered as hard. There was only one item that could be assumed as easy LAS25 (p= .77), discrimination index (D= .27, r= .44) and all distracters of this item were functioning. The remaining 14 items were considered as difficult items, ranges between .12 and .35. Despite their high difficulty, LAP7 (D= .31, *r*= .41), LP29 (D= .33, *r*= .45), and AP4 (D= .30, *r*= .40) were considered as reasonable and needed minor revisions. On the other hand, LS16 (D= .24, *r*= .37), MS8 (D= .22, *r*= .33), AVP26 (D= .18, *r*= .32), and AP21 (D= .16, *r*= .30) were needed to revise marginally in terms of item discriminations.

Finally, there were too difficult items in the test to be considered critically. Items VP1 (*p*= .15), AS13 (*p*= .19), VS9 (*p*= .17), LAP14 (*p*= .15), LS11 (*p*= .12), and VS27 (*p*= .16) were difficult items. These difficult items should be revised dramatically if they would retain for further test implementation.

To sum up, based on the item analysis results, there were 3 items (LAP7, LP29, AP4) that could be retained in test as they were. On the other hand, 3 items (MS8, LAS25, LS16) needed minor revisions, and 5 items (VP1, AS13, AP21, AVP26, VS27) should be revised marginally, and 5 items (LS6, VS9, LAP14, LS11, MS20) should be eliminated for the next administrations.

Summary of Item Analyses

After considering the item analysis results of both Test1 and Test 2, 13 items from Test 1 (LS2, AP3, AP15, AP4, AS19, AP5, AVP26, MS8, VS27, VS18, AP10, LS12, VS24) and 11 items from Test 2 (VP1, AS13, AP21, AVP26, VS27, LAP7, LP29, MS8, LAS25, LS16, AP4) were retained for further administrations. There were totally 20 items, since there were 4 common in retaining items.

4.3 Results of Field Testing III

According to item analysis of second round, 20 items were selected for third administration. For the third round of the field-testing, relatively good functioning 20 items were used for data collection. The purpose of the third administration was to understand how selected items worked together. Similar to second administration both Rasch analysis and Item analysis were conducted. Results of these analyses were used to make final revisions for the final administration.

4.3.1 Results of Rasch Analyses

4.3.1.1 Unidimensionality of Test 3

Through use of fit statistics generated by Rasch analysis, unidimensionality of Test 3 was assessed. The following Figure 4.9 summarizes the unidimensionality check for Test 3. Except one item (AP15) all of the items had infit/outfit statistics within the acceptable range of 0.7 and 1.3, indicating that the test measured a unidimensional construct. In addition, item difficulties ranged from -2.16 (AP15) to 1.69 (VS18). Figure 4.10 also provides a visual display of findings.

ITEMS STATISTICS: MISFIT ORDER

ENTRY NUMBR	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBIS CORR.	ITEMS
8	31	97	.03	.22	1.09	1.0	1.20	1.4	A-.01	AVP26
4	23	96	.44	.24	1.13	1.0	1.19	1.0	B-.08	AP4
14	34	95	-.16	.21	1.08	1.0	1.09	.8	C .00	LAP7
11	25	96	.33	.23	1.09	.7	1.03	.2	D-.01	VS24
19	34	80	-.38	.23	1.06	.7	1.08	.7	E .05	LP29
2	73	97	-1.96	.24	1.02	.2	1.07	.4	F .07	AP3
7	42	97	-.49	.21	1.05	.7	1.04	.4	G .05	MS8
16	15	95	.97	.28	1.00	.0	1.02	.1	H .09	VS27
9	41	96	-.45	.21	1.00	.0	1.02	.2	I .14	VP1
17	22	94	.48	.24	1.02	.1	.94	-.3	J .11	LS16
1	42	97	-.49	.21	1.01	.2	1.01	.1	j .12	LS2
6	27	97	.23	.23	.99	.0	.92	-.5	i .13	AP5
13	13	95	1.16	.29	.95	-.2	.99	.0	h .14	AP13
20	6	59	1.69	.42	.98	.0	.76	-.6	g .14	VS18
12	19	96	.70	.25	.98	-.1	.88	-.6	f .16	AP21
10	44	96	-.59	.21	.96	-.5	.96	-.5	e .18	LS12
18	10	88	1.40	.33	.92	-.3	.80	-.6	d .22	AP10
15	59	94	-1.33	.21	.90	-1.3	.90	-.9	c .29	LAS25
5	21	97	.58	.24	.89	-.8	.75	-1.3	b .29	AS19
3	75	95	-2.16	.25	.84	-1.1	.67	-1.8	a .38	AP15
MEAN	33.	93.	.00	.25	1.00	.1	.97	-.1		
S.D.	19.	9.	1.00	.05	.07	.7	.14	.8		

Figure 4.9 Item Measure Information of Test 3

```
                           ITEMS FIT GRAPH:  MISFIT ORDER

+--------------------------------------------------------------------+
|ENTRY| MEASURE |  INFIT MEAN-SQUARE  | OUTFIT MEAN-SQUARE  |        |
|NUMBR| -     + |0     0.7 1 1.3     2|0     0.7 1 1.3     2| ITEMS  |
|-----+---------+---------------------+---------------------+--------|
|    4|     *   |       :  .* :       |A      :  . *:       | AP4    |
|    8|     *   |       :  *  :       |B      :  .* :       | AVP26  |
|   14|    *    |       :  *  :       |C      :  *  :       | LAP7   |
|   11|     *   |       :  *  :       |D      :  *  :       | VS24   |
|   19|    *    |       :  *  :       |E      :  *  :       | LP19   |
|    2|*        |       :  *  :       |F      :  *  :       | AP3    |
|    7|   *     |       :  *  :       |G      :  *  :       | MS8    |
|    9|   *     |       :  *  :       |H      :  *  :       | VP1    |
|   17|     *   |       :  *  :       |I      : *.  :       | LS16   |
|   16|      *  |       :  *  :       |J      :  *  :       | VS27   |
|    1|   *     |       :  *  :       |j      :  *  :       | LS2    |
|   12|      *  |       :  *  :       |i      : *.  :       | AP21   |
|    6|     *   |       : *.  :       |h      : *.  :       | AP5    |
|   20|        *|       : *.  :       |g      *  .  :       | VS18   |
|   13|       * |       : *.  :       |f      : *.  :       | AP13   |
|   18|       * |       : *.  :       |e      : *.  :       | AP10   |
|   10|   *     |       : *.  :       |d      : *.  :       | LS12   |
|   15|  *      |       :* .  :       |c      :* .  :       | LAS25  |
|    5|      *  |       :* .  :       |b      *  .  :       | AS19   |
|    3|*        |       :* .  :       |a     *:  .  :       | AP15   |
+--------------------------------------------------------------------+
```

Figure 4.10Item Fit Information for Test 3

4.3.1.2 Item Difficulty of Test 3

The following figure provides the variable maps for the Round Three results. Figure 4.11 presents the distribution map of the items and persons, which gave idea about whether the calibration process provided useful information. As seen in Figure 4.3, all items were evenly distributed over (-3, +3 range. On the other hand, the distribution of person was between (-4, +1). Overall the results of the variable maps indicated that the items locations were higher than the person locations, which means that Test 3 was also difficult for the examinees. This observation also confirmed the previous item analysis results from the Round 2.

VS18 in the Test 3 was at the top, whereas AP15 was at the bottom of the vertical line. VS18 was the most difficult item in Test 3 and AP15 was the easiest item in the Test 3. The rest of the items evenly distributed along the vertical line. The distribution of the item locations generally extended along the continuum of person locations. This spread of item locations is desirable as it allows for differentiation among participants through the use of item difficulties.

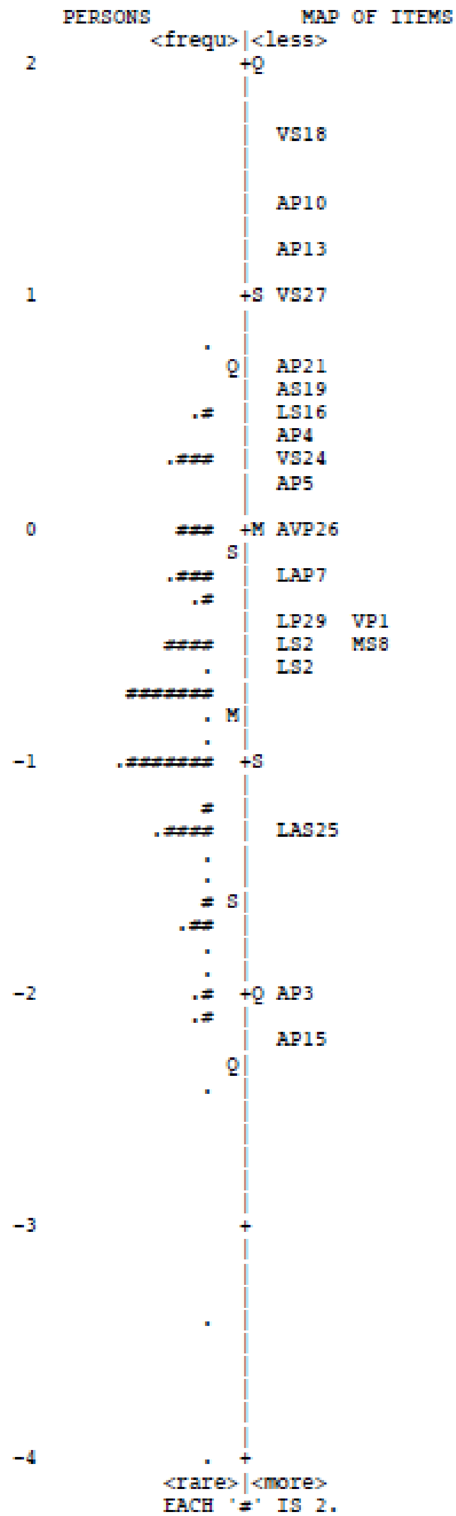


Figure 4.11The distribution map of items and persons of Test 3

The mean of the test scores was 7.32 (SD = 2.36) out of 20 and median was 7.0. Low mean score indicated that test items were difficult for participants. As seen in Figure 4.3, the scores were (Skewness = .167, and Kurtosis = -.551)normally distributed.

4.3.1.3 Reliability and Separation Indices of Test 3

After eliminating the extreme scores (zero and perfect scores), So, Rasch Analysis conducted with 98 non-extreme persons. The following figure summarized the item separation indices.

SUMMARY OF 20 MEASURED ITEMS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	32.8	52.3	.00	.25	1.00	.1	.97	-.1
S.D.	18.8	5.0	1.00	.05	.07	.7	.14	.8
MAX.	75.0	55.0	1.69	.42	1.13	1.0	1.20	1.4
MIN.	6.0	34.0	-2.16	.21	.84	-1.3	.67	-1.8

REAL RMSE	.25	ADJ.SD	.97	SEPARATION	3.80	ITEM RELIABILITY	.94
MODEL RMSE	.25	ADJ.SD	.97	SEPARATION	3.85	ITEM RELIABILITY	.94
S.E. OF ITEM MEAN	.23						

Figure 4.12 Summary of Item and Person Information of Test 3

Findings indicated that there was a good item separation as item reliability is higher than .90. Moreover, items in the Test 1 were separated into almost three difficulty levels (with a Separation of 3.80). The mean and standard deviation of the Infit and Outfit for items indicated that these values were very close to expected values. As seen in Figure 4.4, mean square value of Infit is 1.00 while mean square value of Outfit is .97. Overall, these measures indicated that the items in the Test 3 fit the model well.

4.3.2 Results of Classical Test Theory Analyses- Item Analysis of Test 3

Similar to previous round, at this step participants were quite confident about their responses. The mean of their self-evaluation of correctness of was answers above the neutral value (M= 3.45, Min=2.26, Max=4.32).

4.3.2.1 Item Analysis of Test 3

Item analysis was conducted based on test scores of 99 participants on Test 3, which consisted 20 multiple-choice items. There were three type of information summarized; item difficulty indices, item discrimination indices, and percentage of examinees selecting each response in the Table 4.4.

Item Discrimination Indices

Based on the Ebel's criteria discrimination indices of LS2, AP15, VP1, LAP7, LAS25, and LS16 could be assumed "very good" items in terms of item discrimination because their discrimination indices (D values) were higher than .40. Items AP3, AS19, AP5, MS8, AVP26, LS12, VS24, AP21, VS27, and LP29 could be assumed good in terms of items discrimination, since their discrimination indices were about .30. Items AP4, AS13, and AP10 were marginal items to discriminate the participants of this sample. The last one item VS18 was very poor items in terms of item discrimination. This item was the most difficult item in the test, as expected it had low discrimination indices.

On the other hand, the minimum acceptable point biserial value (Crocker & Algina, 1986) for this sample of 99 is .202. Therefore, items with point biserial coefficients greater than .202 may be retained if other predetermined item criteria are satisfactorily met. Only one item (VS18) was out of values between .202 and .30.Based on this point biserial criteria for the discrimination there were 16 items except items (AP4, AS13, AP10 and VS18) were higher than both discrimination values. Such items would be in the test without any improvement. There three items (AP4, AS13, and AP10) which had an acceptable point biserial value but low discrimination index D. For this case point biserial values were preferred for item analysis. Thus, items AP4, AS13, and LP29 were decided to be functional enough to retain in the test.

According to the results of distractor statistics (See Table 4.4),item analysis indicated that all distracters of all items were good functioning (See Table 4.4). However, one of distracters of items AP4, AS19, VS24, AP21, AS13, LAP7, LS16, and

AP10 functioned more than others. This observation suggested that the reasons why these distracters functioned this much should be investigated in detail.

Table 4.4 Item Analysis Results from 99 Examinees on 20 Item Test 3 (T3)

Item Statistics				Alternative Statistics (%)					
Item	*p*	*D*	*r*	A	B	C	D	E	Omit
LS2	.42	.43	.51	.42+	.19	.12	.18	.06	.04
AP3	.73	.31	.28	.22	.01	.00	.73+	.02	.02
AP15	.73	.52	.46	.00	.00	.05	.06	.73+	.16
AP4	.20	.23	.23	.38	.03	.20+	.09	.22	.09
AS19	.28	.34	.38	.08	.28+	.50	.03	.07	.05
AP5	.32	.29	.25	.32+	.01	.07	.28	.30	.03
MS8	.48	.34	.37	.06	.24	.18	.48+	.01	.04
AVP26	.25	.29	.25	.28	.36	.04	.25+	.05	.03
VP1	.37	.56	.48	.09	.13	.05	.34	.37+	.03
LS12	.55	.29	.23	.08	.15	.06	.11	.55+	.05
VS24	.22	.30	.32	.16	.11	.22+	.25	.10	.17
AP21	.17	.30	.26	.15	.46	.17+	.02	.18	.03
AS13	.21	.23	.24	.52	.04	.21+	.14	.06	.03
LAP7	.46	.42	.52	.46+	.03	.04	.41	.03	.04
LAS25	.75	.45	.43	.11	.02	.02	.75+	.05	.05
VS27	.15	.37	.42	.18	.18	.24	.15+	.19	.07
LS16	.20	.47	.41	.11	.16	.20+	.31	.07	.16
AP10	.14	.25	.43	.44	.14+	.17	.16	.04	.06
LP29	.43	.32	.27	.26	.02	.10	.15	.43+	.05
VS18	.11	.15	.15	.19	.12	.38	.11+	.13	.08

+: The keyed response

Item Difficulty

Considering item difficulty (p) values in ITEMAN results, indices of all of items were normally distributed ranging from (*p*= .11) to (*p*=. 73). Although there was a normal distribution (Skewness = .784, and Kurtosis = -.486) among difficulty levels, then the mean difficulty of test was .36. This can be interpreted that Test 3 was hard for examinees. There were only 4 items of which difficulty levels were considered as ideal, which is about or slightly above .5. Items LS2 (*p*= .42), MS8 (*p*= .48), LS12 (*p*= .55), LAP7 (*p*= .46), LP29 (*p*= .43) were assumed as not hard compared to rest of the test. Based on their D and r values, which were above the critical value, these items were relatively well constructed compared to other items. There were only three items that could be assumed as easy AP3 (*p*= .73, D= .31), AP15 (*p*= .73, D= .52) and LAS25 (*p*= .75, D= .45). These items were good enough and all distracters of this item were good functioning.

Based on the criterion of Haladyna (2004), the rest 12 items were considered as difficult items, ranges between .11 and .37. Despite their high difficulty, items VP1 (D= .56, *r*= .48) and LS16 (D= .47, *r*= .41) were functioning well in terms of their discrimination indices. On the other hand, AS19 (D= .34, *r*= .38), AP5 (D= .29, *r*= .25), AVP26 (D= .29, *r*= .25), VS24 (D= .30, *r*= .32), AP21 (D= .30, *r*= .26), and VS27 (D= .37, *r*= .42) were considered as reasonable. Finally, there were too difficult items in the test to be considered critically. Items AP4 (*p*= .20), AS13 (*p*= .21), AP10 (*p*= .14), and VS18 (*p*= .11) were difficult items. One of them VS18 (*p*= .11, D= .15) should be eliminated considering item difficulty and other discrimination indices. The rest of these 3 difficult items should be revised dramatically if they would retain for further test implementations.

Summary

For the final administration of the test, the researcher and two experts decided that 15 items were enough to cover the intended knowledge domain, which was mathematical knowledge for teaching measurement concepts, specifically, length, area and volume. For this purpose, the difficult items (VS24 and VS18) and low discriminating item LAP7 were eliminated. Despite acceptable statistics, the itemLS12was also eliminated for the final administration. This item was directly related to specialized content knowledge of length measurement. Since the test was designed for the purpose to assess pre-service mathematics teachers' measurement knowledge for teaching, it was decided to try items related to teaching contexts one more time.

4.4 Results of Field Testing IV

According to item analysis statistics of third round 15 items were selected for fourth and the last administration. Analyses were conducted based on test scores of 168 participants on Test 4, which consisted 15 multiple-choice items. Similar to previous administrations both Rasch analysis and Item analysis were conducted.

4.4.1 Results of Rasch Analysis

4.4.1.1 Unidimensionality of Test 4

The following Figure 4.13summarized the item measures of Test 4, as seen in the figure all items were in the intended [.7-1.3] infit and outfit range.

ITEMS STATISTICS: MEASURE ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	SCORE CORR.	ITEMS
5	10	167	2.39	.31	1.00	.0	1.26	.6	.15	AS19
13	24	165	1.43	.21	1.04	.2	1.10	.4	.19	AP10
7	38	165	.86	.18	1.11	1.0	1.19	1.1	.17	AVP26
11	40	162	.77	.18	.97	-.3	.96	-.3	.32	VS27
12	38	138	.62	.19	1.01	.1	1.01	.1	.28	LS16
6	53	167	.39	.16	.98	-.3	1.01	.1	.34	AS13
8	53	162	.36	.17	.88	-1.7	.80	-1.8	.45	AP21
14	66	163	-.03	.16	1.04	.7	1.04	.5	.30	LP29
4	69	165	-.05	.16	.99	-.3	.96	-.5	.36	AP4
1	69	166	-.06	.16	.87	-2.3	.83	-2.1	.48	LS2
9	70	167	-.07	.16	1.09	1.5	1.11	1.2	.26	VP1
15	88	166	-.55	.16	.96	-.7	.97	-.4	.40	MS8
3	102	167	-.91	.16	1.08	1.2	1.16	1.8	.28	AP4
10	145	166	-2.49	.23	.99	.0	.82	-.8	.39	LAS25
2	126	140	-2.67	.28	.95	-.3	.97	-.1	.39	AP15
MEAN	66.	162.	.00	.19	1.00	-.1	1.01	.0		
S.D.	36.	9.	1.27	.05	.06	1.0	.13	1.0		

Figure 4.13 Item Fit Information for Test 4

4.4.1.2 Item Person Map of Test 4

The following figure provided the variable maps for the Test 4 results. As seen in the Figure 4.13, all items were evenly distributed over (-3, +3 range. On the other hand, the distribution of person was between (-4, +2).Overall the results of the variable maps indicated that similar to previous administrations, the items locations were higher than the person locations. That is, item difficulties are beyond the participants' abilities.

As seen in the Figure 4.14, AS19 was the most difficult item in Test 4 and AP15 was the easiest item in the Test 4.

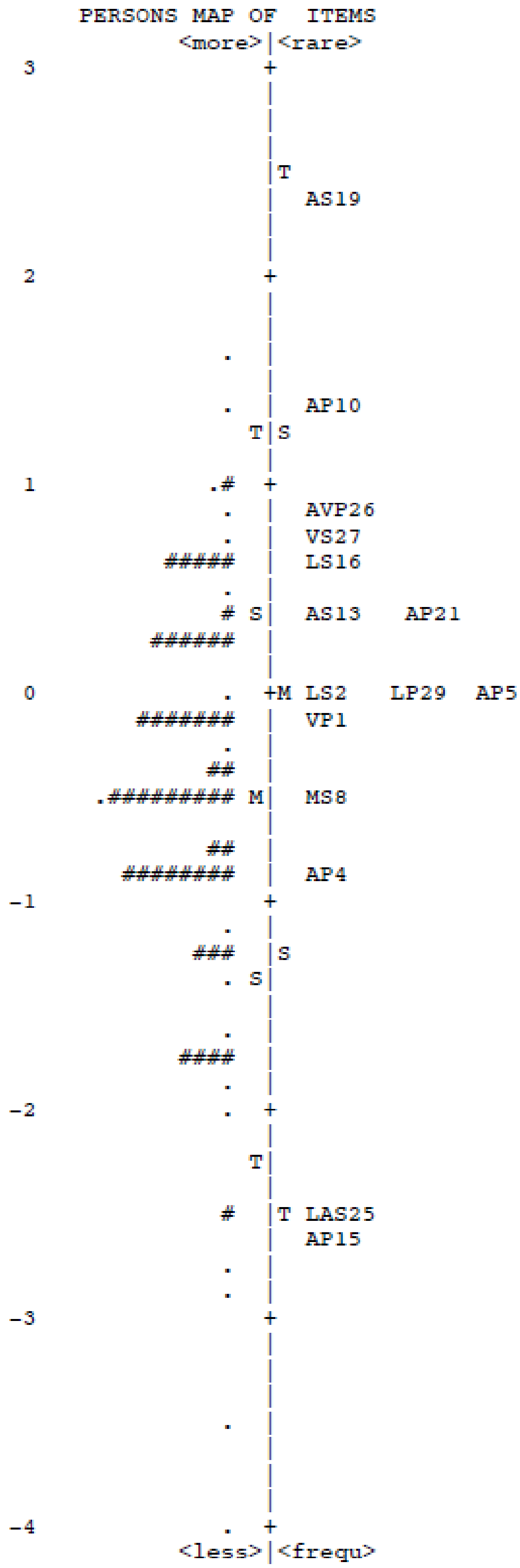


Figure 4.14 The distribution map of items and persons of Test 4

The total test score distribution of participants on Test 4 was displayed in Figure 4.14. Also as seen in the figure the scores were normally distributed (Skewness = -.132, and Kurtosis = -.405). The mean of the test was 5.90 (Std. Deviation=2.215) and median was 6.0.

4.4.1.3 Reliability and Separation Indices of Test 4

After excluding the extreme scores the Rasch analysis was conducted with 167 non-extreme persons. Figure 4.15 summarized the item separation indices.

SUMMARY OF 15 MEASURED ITEMS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	66.1	161.7	.00	.19	1.00	-.1	1.01	.0
S.D.	35.7	9.1	1.27	.05	.06	1.0	.13	1.0
MAX.	145.0	167.0	2.39	.31	1.11	1.5	1.26	1.8
MIN.	10.0	138.0	-2.67	.16	.87	-2.3	.80	-2.1

REAL RMSE .20 ADJ.SD 1.26 SEPARATION 6.36 ITEM RELIABILITY .98
MODEL RMSE .20 ADJ.SD 1.26 SEPARATION 6.42 ITEM RELIABILITY .98
S.E. OF ITEM MEAN = .34

UMEAN=.000 USCALE=1.000

Figure 4.15 Summary of Item and Person Information of Test 4

Parallel to previous findings there was good item separation as item reliability is higher than .90. Moreover, summary in Figure 4.15 gave information about the separation values of items. Items in the Test 1 were separated into almost six difficulty levels (with a separation of 6.36). These measures indicated that the items in the Test 4 fit the model well.

4.4.2 Results of Classical Test Theory Analyses- Item Analysis of Test 4

Item analysis was conducted based on test scores of 168 participants on Test 4, which consisted 15 multiple-choice items. There were presented item difficulty indices, two types of item discrimination indices, and percentage of examinees selecting each response in the Table 4.5.

Table 4.5 Item Analysis Results from 168 Examinees on 15 Item Test 4 (T4)

Item Statistics				Alternative Statistics (%)					
Item	*p*	*D*	*r*	A	B	C	D	E	Omit
LS2	.42	.57	.50	.42+	.22	.11	.14	.08	.02
AP15	.84	.45	.55	.00	.00	.01	.02	.84+	.14
AP4	.62	.35	.28	.06	.06	.62+	.14	.12	.01
AP5	.42	.48	.45	.42+	.00	.02	.03	.50	.02
AS19	.28	.45	.44	.06	.28+	.66	.00	.00	.01
AS13	.33	.30	.29	.48	.02	.33+	.14	.02	.01
AVP26	.20	.21	.19	.18	.53	.02	.20+	.04	.02
AP21	.35	.51	.47	.23	.01	.35+	.26	.12	.03
VP1	.42	.26	.28	.09	.06	.28	.42+	.14	.01
LAS25	.96	.13	.43	.01	.00	.01	.96+	.01	.02
VS27	.26	.28	.32	.15	.17	.35	.26+	.03	.04
LS16	.23	.30	.31	.14	.12	.23+	.23	.06	.21
AP10	.14	.16	.21	.39	.14+	.22	.19	.03	.02
LP29	.42	.20	.24	.12	.01	.18	.24	.42+	.03
MS8	.56	.43	.35	.06	.12	.21	.56+	.04	.01

+: The keyed response

Item Discrimination Indices

Based on the Ebel’s criteria discrimination indices of LS2, AP15, AP5, AS19, AP21, and MS8 could be assumed “very good” items in terms of item discrimination because their discrimination indices (D values) were higher than .40. Items AP4, AS13, and LS16 could be assumed good in terms of items discrimination, since their discrimination indices were either equal or higher than .30. On the other hand, items VP1 and VS27 could be considered as good, since their discrimination indices were about .30.

Items AVP26, LAS25, AP10 and LP29 were marginal items to discriminate the participants of this sample. Items LAS25 and AP10 were very poor in terms of item discrimination. The item LAS25 was the easiest and the items AP10 was the most difficult item in the test, as expected they had low discrimination indices.

The minimum acceptable point biserial value for this sample of 168 is .154 (Crocker & Algina, 1986). Therefore, items with point biserial coefficients greater than .154 may be retained if other predetermined item criteria are satisfactorily met. All items were found to have point biserial indices above .154, five items (AP4, AS13, AVP26, AP10, LP29) had correlational indices between .154 and .30. These items (AVP26 and AP10) along with the items would be flagged for unsatisfactorily low p-values.

Based on this point biserial criteria for the discrimination there were 11 items except items (AVP26, LAS25, AP10 and LP29) were higher than both discrimination values. Such items would be in the test without any improvement. Those four items had acceptable point biserial indices but low discrimination indices D.

According to the results of alternatives statistics (See Table 4.5),item analysis indicated that the alternative statistics of items indicated that all distracters of all items were good functioning (See Table 4.5). However, one of distracters of items AP5, AS19, VS27, LS16, and AP10 functioned more than others. This observation suggested that the reasons why these distracters functioned this much should be considered critically.

Item Difficulty

Considering item difficulty (p) values in ITEMAN results, indices of all of items were normally distributed ranging from (p= .14) to (p=. 96). Although there was a normal distribution (Skewness = 1.15, and Kurtosis = .926) among difficulty levels, then the mean difficulty of test was .43. This can be interpreted that Test 4 was not very hard for examinees. Besides, there were only 4 items of which difficulty levels were considered as good, which is about .5. Items LS2 (p= .42), AP4 (p= .62), AP5 (p= .42), VP1 (p= .42), MS8 (p= .56) were assumed as not hard compared to rest of the test. Based on their D and r values, which were above the critical value, these items were relatively well constructed compared to other items. There were only two items that could be assumed as easy AP15 (p= .84, D= .45), LAS25 (p= .96, D= .13). Based on Haladyna (2004) criterion, the rest 7 items were considered as difficult items, ranges between .14 and .35. Despite their high difficulty, items AS19 (D= .45, r= .44) and AP21 (D= .51, r= .47) were functioning well in terms of their discrimination indices. On the other hand, AS13 (D= .30, *r*= .29) and LS16 (D= .30, *r*= .31) were considered as reasonable. Finally, there were too difficult items in the test to be considered critically. Items AVP26 (*p*= .20), VS27 (*p*= .23), and AP10 (*p*= .14) were difficult items.

4.5 Further Validation Evidences

4.5.1 Raw Scores and Rasch measures

The relationship between total scores and Rasch person measures was investigated using Pearson product-moment correlation coefficient. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity and homoscedasticity. There was a strong, positive correlation between the two variables [r=.946, n=144, p=.00], with total scores of test and Rasch measures of persons.

4.5.2 GPA and Raw Scores

The relationship between total scores and GPA grades of pre-service teachers was investigated using Pearson product-moment correlation coefficient. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity and homoscedasticity. There was a moderate positive correlation between the two variables [r=.38, n=141, p=.00], with total scores of test and GPA grades of persons.

4.5.3 GPA and Rasch Measures

The relationship between Rasch measures and GPA grades of pre-service teachers was investigated using Pearson product-moment correlation coefficient. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity and homoscedasticity. There was a moderate positive correlation between the two variables [r=.40, n=141, p=.00], with Rasch measures and GPA grades of persons.

CHAPTER 5

DISCUSSION AND CONCLUSION

The instrument developed in this study was a multiple-choice instrument, which seeks to assess conceptual understanding of the pre-service elementary mathematics teachers' mathematical knowledge of measurement concepts for teaching, specifically on the concepts of length, area and volume. The current study was distinctive in mathematics education literature, since it was one of the preliminary studies on test development in which the mathematics knowledge for teaching was considered, especially for the case of Turkey.

As explained before, related literature suggests that there is a need for presentation of the procedure of developing valid and reliable measures for teachers' knowledge for teaching (Hill, et.al., 2004, Hill, et. al., 2007).The product of this study was an instrument that could assess pre-service teachers' mathematical knowledge for teaching (MKT) on the concepts of length, area, and volume measurement. Thus, it was aimed to contribute to fill the gaps for lack of valid measures to be used for assessing elementary mathematics pre-service teachers' MKT. Findings on the items were intended to provide a contribution of this study to the determination of teacher qualifications. Therefore, this research intended to contribute to the body of research those curriculum developers, educators, faculty member, and bureaucrats can utilize in teacher education programs.

The main significance of this study was that it involved the development of one of the prototypes for valid and reliable measures for the assessment of MKT, as this study was conducted to fill such a gap in the literature. Thus, the design of this study included producing one of the examples of these necessary measures. Describing a methodology for developing such an instrument and creating survey items that can be used as a basis in future tools for assessing teachers' PCK were the other purposes of the study in line with the need mentioned. For this purpose, the results from the Chapter IV will be summarized and discussed: the performance of the test items and validity of the instrument was a major issue to be discussed in this section. The chapter will be closed with a presentation of limitations, implications and recommendations, and conclusions for further research.

5.1 Item Construction andRelationship of theResultstoPreviousResearch

The model of Mathematical Knowledge for Teaching (MKT) was reviewed in detail within the framework of the existing research described in Chapter II. Ball and her colleagues (2008) propose a practice-based theory of content knowledge for teaching built on PCK (Ball, et. al., 2008). The domain map for "Mathematical Knowledge for Teaching" points out the knowledge the teachers are expected to have. The domain map -Mathematical Knowledge for Teaching - has two main subdomains; Subject Matter Knowledge (SMK) and Pedagogical Content Knowledge (PCK) with several theorized knowledge domains (Ball et al., 2008). The current test was modeled after the Learning Mathematics for Teaching (LMT) instruments for assessing in-service teachers' mathematical knowledge for teaching (Hill et al., 2004).However, two main gaps were presented directly related to the test development for the assessment of MKT: the first one was the lack of clear definition of the construct (in other words fuzzy boundaries between subdomains) and the second one was the difficulty of developing items for assessment of teacher knowledge.

As explained in Chapter III, items were constructed to address the portion of the Specialized Content Knowledge (SCK) and Pedagogical Content Knowledge (PCK) domains within the pre-determined learning objectives of measurement concepts in Turkish elementary mathematics program. Based on the MKT model, the construct of

PCK was defined by Knowledge of Content and Students (KCS), Knowledge of Content and Teaching (KCT), and Knowledge of Content and Curriculum (KCC). However, the boundaries between these subdomains are not clear enough to make clear decisions when considering real teaching cases or teaching scenarios. Thus it becomes difficult to construct objective measurement items focusing specifically on these subdomains. For this reason, items of MKT-M measure can be categorized in two clearly distinguishable subdomains: SCK and PCK.

The structure of the items was multiple-choice item format. Although it is not suggested to use complex multiple-choice item format in test development literature (Haladyna, 2004), in this study some items (AP4, VP1) related to teaching context were re-designed as complex multiple-choice item format. Based on the findings of interviews, it was understood that there was no unique approach for the given cases and the solution might sometimes depend on the interaction of teacher and student. Thus, there might exist more than one answer. Moreover, item difficulty and item discrimination indices of these complex multiple-choice items became acceptable after revision of items.

5.2 Scoring of the Test Results

In test development literature Classical Test Theory was one of the main framework to analyze items and to make decisions on them. However, these tools have some limitations as also summarized in Test Development section. Specifically, this type of scoring was dependent on the characteristics of sample as well as the characteristics of items. Thus it becomes difficult to determine standardized scores for each item in different forms of tests. On the other hand, in Rasch analysis, estimation of item and person parameters was carried out through some underlying mathematical patterns. Thus this analysis provides a standardized scale, which provides an opportunity to make more meaningful comparisons between different test forms and between different samples (Hambleton, et. al., 1991). For this reason, although there exist the item analysis from the perspective of Classical Test Theory in this study; the main scoring approach of items of MKT-M measure was Rasch Analysis.

5.3 ReliabilityandValidity

For the test development the vitality of assessment of the reliability and validity of the instrument is inevitable. For this purpose, the reliability of an instrument could be measured by different tools such as: coefficient alpha (from the perspective of classical test theory), coefficient omega (for not strictly unidimensional tests), separation indices and reliabilities (from the perspective of Item Response Theory), or any other.

For the current case, there were two reasons for using separation index and reliability of Rasch Analysis for the reliability measure. The first one was that the instrument was unidimensional in nature. The second one was that the value of coefficient alpha is strictly dependent on sample and item characteristics. Within the Item Response Theory framework the standard error of reliability is not constant, instead it is a function of the standard error of measurement for the test across the ability distribution. Thus it does not provide a single estimate of reliability. The single reliability estimate value is the averaged error across the ability distribution. Thus in order to take the effect of sample and item characteristics under control, item reliability of the Rasch Model was taken into consideration. The reliability value for this study was .98, which is quite high. Moreover this value is not sample-dependent and is not an estimate of single administration.

On the other hand, the value of reliability of Cronbach Alpha was not reported because of the low value of it. There might be varying factors that negatively affected the value. First, the participants were not accustomed to this kind of tests, in terms of neither structure nor content. Moreover, the design of the teaching methods courses in undergraduate teacher education programs was not parallel to the coverage of the test. Accordingly, it was observed that majority of the participants found the items were difficult to reach correct answer during the test administrations. Secondly, the test was designed as an achievement test and participants were not well prepared for the test taking. Another possible explanation for why the Cronbach alpha was low was that the lower number of items in test booklets (about 16 items) might have an effect on the reliability (Crocker &Algina 1986; Nunnally, 1967). Considering these people characteristics, it is suggested to use item response theory reliability estimations for the future implementations of this test. For this reason the reliability value of the test provides more valid conclusions for the Raschlogit scores. Thus it is advised to consider Raschlogit scores for the further implementations of this instrument.

As mentioned before, addressing test validity is a very important step of the test development process. Validity is a complex construct, and there is no single measure to point to when trying to establish the validity of tests, test scores, and most importantly the inferences that are drawn from them. Instead, evidence of validity must be collected from a variety of sources. As a holistic approach the construct validity is considered to encompass all other types of validity (Loevinger, 1957; Messick, 1989). Based on this conceptualization, Loevinger and Messick point out three components of construct validity: substantive, structural and external validity of measures. The specific evidences for these three components of validity will be explained respectively.

Substantive validity comprises the conceptualization of the domain and development of an initial item pool for measures. Conceptualization, literature review and creation of item pool are the main steps should be considered in measure development (Downing, 2006). The substantive validity goes beyond making this claim based on the fact that the items really represent the intended construct. This validity can be derived from a domain space clearly specified in advance and judged by experts to be representatives of the area. For this case, the results of factor analysis did not produce meaningful results. As explained before, the majority of the items were difficult for test takers. Despite lack of Factor Analysis results, the strictly unidimensional structure of the multiple forms of items in Rasch Analysis might be interpreted as all items covering the same construct: mathematical knowledge for teaching.

Structural validity of the measurement is complementary to the substantive validity and mainly encloses the item selection and psychometric evaluation of the test (Messick, 1989). According to Messick's definition, the substantive validity relies heavily on the analysis of test scores. Structural validity issues of test were addressed by conducting individual analysis of keyed and distracter responses. Item analysis included the individual analysis of item difficulty, item discrimination, and correlation coefficients. The values of these coefficients provided evidence whether it was a good item or not in terms of general objective test theory or not. Each step explained in Item analyses and Rasch analyses provided evidence to support the structural validity of the test.

While constructing the instrument MKT-M, the domain was initially specified as measurement. Thistopicwas regarded that it covers important concepts based on the needs of Turkish elementary mathematics education program. Items were constructed based on the conceptual nature of the topics rather than recall and computation. The test has been implemented several times and interviews with pre-service teachers have been used to discuss the functionality of individual items. Based on the interview findings, malfunctioning items have been eliminated or rewritten.

In addition, item and distracter analysis has been conducted for all items to determine item difficulty, discrimination, and the effect on test reliability. Additionally, Rasch Analysis has been conducted and this information has been paired with classical test statistics to identify items, which would benefit from further revision. The portray of the items were not determined only by items parameters or distracter analyses, fit indices of Rasch Analyses but also provided an explanation for the reasons of why there existed problematic items. Administration of both Classical Test Analyses and Rasch Analyses provided an opportunity to determine greater number of problematic items. Although item analyses could be investigated by the use of CCT tools, results of Rasch Analysis provided a better view about the distribution of item difficulty versus person ability. These instruments revealed the functionality of items.

Face and content validity were determined by consulting in-service and pre-service teachers, mathematics educator, test development expert and test developer. The issues of content and construct validity were addressed by the extensive literature review. In order to ensure the substantive validity of the test, four in-service teachers evaluated the reality and appropriateness of cases according to real classroom contexts. Besides in-service teachers, the rest of three experts were involved in evaluating the structure of test. Content validity was also addressed by using the expert panel to ascertain adequate topic coverage. One expert who was specialist in mathematics education not only investigated content of items but also checked the items with respect to the categories of intended domain in the table of specification. One test development specialist analyzed the content and structure of items in terms of test development principles. One test developer investigated the wording and language of items for readability and understandability. Before implementing items, 7 participants were asked to complete the test and make comments and criticisms for face validity of test. To sum up, every attempt to provide evidence in test development steps also supported the substantive validity of the test.

Finally, the *external validity* of how the test interacts with other measures (both test and non-test behaviors) was predicted by construct theory (Messick, 1989). At this point there was no other measure that could be comparable to the scores of MKT-M. The only thing that could be done was the analysis of the correlation between persons' GPA scores and total test scores. There was a moderate positive correlation between the two variables [r=.38, n=141, p=.00], total scores of test and GPA grades of persons. This was not a surprising result since the GPA scores are measure of course passing scores of participants, which is also affected by other components such as attendance, homework, project work etc. Moreover, the content and the structure of the MKT-M items were novel for participants. There may be higher correlations between the mid-term/final scores of specific teaching methods courses, which are more focused measures of teaching mathematics content. Similarly, There was a moderate positive correlation between the two variables [r=.40, n=141, p=.00], with Rasch measures and

GPA grades of persons. As explained before the GPA scores may somehow be a holistic measure of undergraduate education, instead specific teaching content. Thus, external validity of the measure depends on the quality of initial steps of test development. Validation work begins to assess; using non-psychometric methods, whether the items tap the intended construct. Briefly, the validity of each step in test development will ensure the whole validity of the measure. To sum up, assessment of validity is an ongoing activity. The more data are collected, the more revisions are made.

While the theoretical consideration of MTK have led to a better understanding of teacher knowledge and effective teaching and was an important milestone in teacher knowledge literature, there existed several measurement problems, accordingly needs and many innovations. The MTK-M was still new to the assessment of teacher knowledge movement. Although, total scores were low and gains are minimal, outcomes have been consistent with researcher's expectations. It is hoped that with further use and research, the MTK-M will inform efforts to develop instructional strategies for teaching methods courses. This chapter outlined future research goals for the MTK-M and presents some preliminary findings from research in these directions.

5.4 Performance of theMKT-M items

In order to be confident about the functionality of items, especially for determining the poor and ill structured items, item analysis was an important step for test development. In order to make valid interpretations for the Results of RaschAnalysis, the data is required to fit the model reasonably well. For the current study, the final form of all items were unidimensional, in other words all items addressed only one construct. The mean infit and outfit for person and item mean squares were expected to be 1.0. For these data, they were 1.00 and 1.01 respectively. The mean standardized infit and outfit are expected to be 0.0. Here they are -.1 and .0. Based on the cut-off criteria (0.7-1.3) items with a standardized infit standard deviation of 1.0 have acceptable fit overall (Bode & Wright, 1999).

The other overall statistic about the MKT-M test is the index of spread of the item positions. For the present case the observed item separation was 6.36 (real), and it was 6.42 (model) when there is no misfit item in data. Since the value was higher than 1.0 this observation could be interpreted such that the items were positioned along a continuum rather than dichotomous locations. Since the instrument was a kind of achievement test, accordingly different categories for items were intended. Otherwise the test would have only easy or difficult items in a total of two categories. The main factors, which affect separation index for items, are sample size, fit indices and error estimates. When sample size becomes larger, separation index increases and error decreases. The item separation reliability estimate for these data was .98. The mean of the participants' ability scores was -.47 when the mean for items was set as 0.0, which could be interpreted as these items'difficulty for the participants. This observation also confirmed the following results of item analyses from the classical test theory perspective.The difficulty level of items varied from high (Item AP10, $p= .14$) to low (Item LAS25, $p= .96$). The mean and median of the test were M=5.90 (Std. Deviation=2.215, out of 15) and median was 6.0.

Discrimination index of these items, on the other hand, again showed different values, from (Item LAS25) D=.13 to (Item LS2) D=.57. The discrimination index for six items was below .30. This low discrimination indicates that the relationship between total scores of participants and scores on the item was relatively weak. These items somehow failed to discriminate the participants based on their mathematics knowledge for teaching. The other interpretation for this may be that the pre-service teachers might have common misconception or incomplete knowledge on the concept which the items cover.These items should be handled in depth for further research.

Another investigation on the data was conducted by distracter analysis. It was observed that although all distracters of items were functioning quite well, some distracters functioned at a higher degree than expected. Based on the interview findings some of these cases might be the signal of common misconception of pre-service teachers.

This might negatively affect item discrimination (Haladyna& Downing, 1993), which might result in the lower precision to measure mathematics knowledge for teaching of participants.

For example, in the context of Round I, PMTs were asked to think aloud on how and why they chose certain answers for each item. Although the main aim of these interviews was to revise the item, this analysis provided an opportunity to get idea about how items were interpreted by the test takers as well as the structure of items and functionality of distracters. Since the content and structures of some items were revised after interviews, these interview findings were not reported in Chapter IV in detail. However, these findings provided some suggestions why some items fail to function as expected despite iterative revisions. At this point quotations from interviews were used to illustrate participants' reasoning for items.

For the item AP10, PMTs were asked to determine the prerequisites of the surface area of cone (See Appendix A).The item statistics of AP10 based on the Round IV was below the expected values (p=.14, D=.16, r=.21).The analysis of open-ended responses and findings of interviews signified that PMTs had some problems related to surface area of cone.

To begin with, some interviewees indicated that the concept of the cone had been problematic for themselves since their schooling years.

#ID12: The cone has been the most untasteful concept for me since my schooling period. It feels complicated and difficult. I still do not know the formula of the surface area [of it].

In fact one participant stated that he even had no idea about how to calculate the surface area of cone. Hence AP10 was one of the difficult items in the test for him.

Furthermore, majority of interviewees experienced a difficulty to construct the net of the cone. Many of them claimed that the surface area of cone consists of only the circle sector. When the participants were asked whether they remember the formula of the surface area of the cone, surprisingly many of them answered with the formula of the volume of the cone. Interview findings in Table 5.1 might suggest some reasons why this item functioned this way.

Table 5.1 Frequencydistribution of PMTs' justifications on Item AP10

Alternatives (%*)		Main categoriesforPMT'sreasoning on distracters
A	(39%)	Area of circle, area of circlesector, circumference of thecircle, andheight of conearenecessaryforsurfacearea of cone [Acceptingallalternativeswithoutcriticizing.]
B	(14%)	[Correctanswer] Area of circle, area of circlesector, andcircumference of thecirclearenecessaryforsurfacearea of cone.
C	(22%)	Area of circleandarea of circlesectorarenecessaryforsurfacearea of cone. [Ignoringthecircumference in theprocedure of areacalculation]
D	(19%)	Area of circleandheight of conearenecessaryforsurfacearea of cone. [Confusingsurfaceareaandvolume]
E	(3%)	Onlyarea of circlesector is necessaryforsurfacearea of cone.

*: These percentages were retrieved from item and alternative statistics of Round IV

5.5 Limitations of theCurrentResearch

Like any other research study, this study also had limitations that should be taken into account, such as: focusing on restricted mathematical concepts –measurement-, inexperienced participants, using unfamiliar test content and structure for the participants, and using convenient sample for the last two rounds.

The first limitation for this study was focusing on just a tiny part of mathematics concepts. The instrument covered only the measurement concepts of length, area and volume, disregarding all other subjects in mathematics education. Thus, it was difficult to generalize the results to all other teaching domains of mathematics.

The other limitation was that the participants were unfamiliar to the test structure since the participants had never encountered such a test when the test was administered. Furthermore, the content of the teaching courses was not parallel to the content of the test. Thus, it becomes difficult to make clear interpretations about the item statistics.

The last but not the least limitation was about the procedure of data collection. Although there was a random sampling method for the second round, for the last two round, data was collected the universities in Ankara. So, using such a convenient sample for the last administration was an important limitation to generalize the results.

5.6 Implications for Practice

In the context of the current study aninstrument was developed. Based on the results of the current study,this instrument can measure pre-service teachers' mathematical knowledge for teaching to some extent. This means that PCK, which was explained in an entirely theoretical way in the studies of Shulman (1986, 1987), has practical implications in real contexts. The initial signs of analyses results will motivate for the similar studies to be performed in future and for the development of more tests in this field. Besides, this will trigger additional research questions for measuring teacher competence by transferring the teaching knowledge from the framework of the education of teachers to another dimension; measurement of teacher knowledge. By this way, it will be possible to take remarkable steps in the field of accurate identification of teaching knowledge.

As stated earlier, the main aim of study was to develop valid and reliable multiple-choice items on specific mathematics concepts for measuring the MKT of pre-service teachers. During the item development process, it was observed that pre-service teachers had some problems, misconceptions and difficulties related to these specific mathematical concepts. They faced difficulties about conceptualizing the measurement, making connections betweenmeasurement and other mathematicalideas, and attributing meanings to mathematical formula used for the concept. However, revealing pre-service teachers' misconceptions and difficulties was out of the scope of the current study. Thus, in-depth data collection and data analysis on this issue were not conducted and the findings reached at the end of this study remained at observation level and were reported in the form of brief notes in the relevant sections of the Chapter 4. Thus, these observations present a few suggestions to mathematics teachers, in terms of mathematics teaching and teacher education. Based on these observations, a few suggestions for mathematics education can be listed as belowthrough main aspects of PCKpointed by Shulman (1986, 1987).

Shulman (1986) emphasized two major dimensions related to PCK. First, he mentioned that PCK included “the most useful forms of representation of ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations - in a word, the ways of representing and formulating the subject that make it comprehensible to others” (s.9). It means that for a person to deliver an idea to another one in a functional form, first, the person needs to have clear and comprehensive ideas on the related concept. In other words, the first step is that teachers are required to make the mathematics comprehensible to themselves. Then, they can make the subject matter comprehensible for others especially for students. Otherwise, it would be unfair to expect the teachers to teach the mathematical concepts that they have not yet fully conceived, in a meaningful way.

However, it is even among the literature thatpre-service teachers and even in-service teachers may have similar difficulties and possess similar misconceptions with the students (eg. Even, 1990;van Driel, 1998, Ward, 2004).In the context of this study, it was observed that some pre-service teachers also experienceddifficulties in the same way with the student cases given in the instrument items. Naturally, they were unable to identify the student difficulty that they came across with. Thus they could not sense the main reason of the student’s underlying difficulty. Hence their performances remained poor at item cases given in the instrument, where they were expected to make a decision and to intervene as a teacher. Further observation made during the interviews at the time of item development processwas that, the pre-service teachers expressed that they did not have much experience about the thinking approaches of students, and that they had not questioned their knowledge about this in their courses. Consequently, they experienced difficulties in higher-level questions, in which they were expected to identify the thinking of the student and intervene to the given situation as a teacher, rather than only and directly utilizing mathematical knowledge. Following all these observations, it was understood from the analyses made on the developed Test of

Mathematical Knowledge for Teaching Measurement (TMK-M) that the items, which coveredSCK, performed better in terms of difficulty and discrimination and that the indices of difficulty and discrimination were lower than expected for the items that were prepared for measuring PCK.

Shulman's second point (1986) was that "pedagogical content knowledge also includes an understanding of related about what makes the learning of specific topics easy or difficult: the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning of those most frequently taught topics and lessons." (p.9).There are also empirical studies in literature, which highlights the importance of the teachers' knowledge on the students for meaningful learning (eg. Carpenter, et al., 1992; Ryan & Williams, 2003). According to these studies the more knowledge of the common mathematical errors and misconceptions of students that teacher have, the more understanding and comprehension of student cases, dependently the more greater vision of teaching and learning.

The first and the main suggestion of the currentis towards strengthening the mathematical knowledge, which form the basis for the PCK of the pre-service teachers. In order that the pre-service teachers should overcome their difficulties and misconceptions about fundamental mathematical ideas which they have brought from their schooling years before becoming teachers.These concepts need to be discussed and those problems need to be solved during their undergraduateeducation.The pre-service teachers should be provided opportunities to reflect on mathematical concepts, and a special emphasis should be given to the courses on basic mathematical concepts. If such courses exist, then they should be increased in number. Within the framework of these courses, not only discussions on the underlying ideas of mathematical concepts, but also studies on connecting different concepts, multiple representations, and verification and proof of mathematical ideas should be carried out.

Secondly, the knowledge of pre-service teachers on students' characteristicsneeds to be advanced. In addition to understanding the characteristics of students, the pre-service teachers should have the ability to understand the student in possible cases that they may face, to identify the challenge or misconception that the student is going through and to have an idea about the origin of this situation. To this end, before graduation from the education faculty program, the identified characteristics of students, their tendencies, the challenges and misconceptions they face should be delivered within the context of specific mathematical concepts to the pre-service teachers. Another suggestion is to extend the visions of pre-service teachers in relation to various instructional methods, again with an aim of strengthening the teaching knowledge.

Another implication lies in the teacher education programs throughout different universities. During the sub analyses, it was observed that pre-service teachers in different universities have different answering patterns with respect to test items. However, since revealing the reason for such answering patterns was outside the scope of this study, it was not reported in detail. The generation of these differences between universities may be a sign for difference in the content of method courses given in different universities especially the conduction of method courses with different contents and the lack of standardization. While the data gathering studies, interviews were made with the instructors of method courses, during which it was observed that the course contents designed, the resources used and the approaches showed variation. Because of this, there is a need to put more emphasis on the necessary accreditation criteria identified by the Higher Education Institution, and to perform studies for enabling the pre-service teachers in all education faculties to acquire similar experiences. Additionally, there is a need for conduction of long-term studies in different patterns, in order that the underlying reasons for such answer patterns of pre-service teachers can be scientifically explored.For this aim, it is possible to perform long-term studies by utilizingthe test developed within this study as well as similar tests. In the longitudinal research designs, iterative administrations of the current instrument will provide detailed information about pre-service teachers' knowledge for teaching development and will help monitoring acquisition steps. Thus, the results of such longitudinal studies will also provide information about the instructional programs, course contents to foster mathematical knowledge for teaching of pre-service teachers.

5.7 Implications for Future Research

Related literature emphasizes the necessity of valid and reliable measures to test and to explore the nature and structure of the pedagogical content knowledge, or more narrowly, mathematics knowledge for teaching. Thus, measures peculiar to construct are necessary in order to further explore and validate the construct itself. Test development initiations require reconsidering the theoretical construct itself. Moreover, extent and content of the measures provide an opportunity to re-define dimensions and sub dimensions of the construct, dependently on the structure of theoretical propositions about the construct. Obviously such kind of iteration provides feedback not only for the theoretical verification studies, but also for validation of the measurement cyclically. The current study was aimed to fill such a gap in the literature. Furthermore, the development of the MKT-M items questioned the boundaries of theoretical model- the construct (Ball, et al., 2008) within the framework of test development and practical applications of test development. The subdomains of MKT model were very difficult to separate and accordingly it was very hard to define each subdomainin real contexts. Thus, it was decided to define the construct in two dimensions: SCK and PCK.

At this point, in the literature there are only two points on which researchers provide some consensus. The first one is that experience has a great effect on the development of PCK, (Gess-Newsome, 1999; Loughran, et al., 2007; Magnusson, et al., 1999; van Driel, Verloop, & de Vos, 1998) which was referred to as knowledge for teaching in the current study. The second one is that subject matter knowledge and general pedagogical knowledge arethe main prerequisites for the development of knowledge for teaching (Magnusson et al., 1999).Except for these two common points, the researchers have not yet reached a common ground on the structure of PCK and its sub domains. For example, SCK and KCS – sub domains of MTK which was assumed to be unidimensional in the studies conducted - have been reported to be not unidimensional and to show a structure including more than one factor (Schilling & Hill, 2007). It can be concluded that the theoretical assumptions of PCK construct may conflict with the research results in some cases. Although this situation causes difficulties for researchers in many fields extending from writing test items to deciding analyses, studies within the present information and assumptions are needed for the structure of PCK to become crystallized and clear. The results of these studies will provide the theory about this concept, while the changes in the theory will provide the evolution of measurement tools. At this point, the basic fact is the existence of a knowledge which separates the teaching profession from the other professions, although its structure is not entirely known. On the other hand, in order for the uncertainties on construct to be removed, it seems that there is no other solution than performing more studies on this issue.

Furthermore, the development of the MKT-M items questioned the boundaries of theoretical model- the construct (Ball, et al., 2008) within the framework of test development and practical applications of test development.Based on this finding, future research can focus on developing items of other domains of MKT separately. Explanatory and Confirmatory factor analyses will be helpful for validation of MKT model. These analyses allow the researchers to test hypothesized model and structure of a measure.

The other implication for the research area is about the determination of analysis model. At this point, it seems psychometric properties of dichotomously scored tests are more advantageous than other kind of tests with respect to validity and reliability. However, this kind of dichotomously scored tests might fail to explore errors, misconceptions, or other thought patterns related with test items. For this reason it may be useful to construct partial credit test scoring models for further research designs to detect underlying ideas. Alexander, Kulikowich, and Schulze (1994) state that responses of teachers as well as responses of students may not be random; indeed those response patterns may be an indication of any kind of misconception or difficulty. For this reason, especially for tests peculiar to complex constructs, such as knowledge for teaching, it may be better to analyze test scores by using polytomousmodels where distracters are appropriately constructed in order to detect underlying thought patterns.

To sum up, being aware of complexity of teaching, it should be realized that using a single method may be limited for assessing such a complex process. Ideally for a teacher assessment tool to be comprehensive and coherent, all of the components of teacher knowledge should be addressed. In short, for the valid and comprehensive assessment of teacher knowledge for teaching, it is required to use multiple data gathering designs for research.Interview findings suggested some underlying reasons such as quantitative reasoning, beliefs about teaching mathematics, and teaching self-efficacy. In order to elicit these relationships, a structural equation modeling (Benson, 1998) approach could also be used.

5.8 Conclusion

The main aim of this study was to develop a valid and reliable test to measure the pre-service mathematics teachers' mathematical knowledge for teaching. For this reason the main focus of this study was to establish construct validity evidence for the newly developed MKT-M instrument in order to make healthy inferences from the scores. However, it can be concluded that there was limited empirical support to make inferences about pre-service teachers' MKT considering the scores of MKT-M items. Although all items were not ideally functioning, most of them signaled some clues to construct better measures for assessing pre-service teachers' MKT. At this point the current study was one of the preliminary attempts to develop an instrument on MKT specifically in Turkey. This study was designed as much comprehensively as possible. The findings of the study presented the reasoning strategies and problem solving approaches of pre-service teachers on multiple-choice items. However, there was no other similar instrument to compare the test results and to make some further inferences.

To sum up, it is believed that the instrument MKT-M will have a contribution to the test development for assessing MKT literature by introducing a scale that could be used to assess pre-service teachers' mathematical knowledge for teaching specifically on the concepts of length, area, and volume measurement. It is also believed that the test development process will illuminate the further test development studies.

REFERENCES

Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). How subject-matter knowledge affects recall and interest. *American Educational Research Journal, 31*, 313–337.

Ball, D L. (1990). Prospective Elementary and Secondary Teachers' Understanding of Division. *Journal for Research in Mathematics Education*, *21* (2), 132-144.

Ball, D. L, & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.),*Multiple perspectives on the teaching and learning of mathematics (*pp. 83-104). Westport, CT: Ablex.

Ball, D. L. (1990). The mathematical understanding that prospective teachers bring to teacher education. *The Elementary School Journal*, *90* (4), 449 (-466.

Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *The Elementary School Journal*, *93* (4), 373-397.

Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator*, *29* (1), 14-22.

Ball, D. L., Lubienski, S., & Mewborn, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.),*Handbook of research on teaching* (4th ed., pp. 433 (-456). New York: Macmillan.

Ball, D. L., & McDiarmid, G. W. (1990). The subject matter preparation of teachers. In W. R. Houston, M. Haberman, & J. Sikula (Eds.),*Handbook of Research on Teacher Education* (pp. 437 (-449). New York: MacMillan.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content Knowledge for Teaching: What Makes It Special? *Journal of Teacher Education*, *59* (5), 389 (-407.

Ball D. L., & Wilson, S. M. (1990).*Knowing the subject and learning to teach it: Examining assumptions about becoming a mathematics teacher*. (Research Report 90-7). East Lansing, MI: National Center for Research on Teacher Education. ED 232 207

Ball, D. L. (1988).*Knowledge and reasoning in mathematical pedagogy: Examining what prospective teachers bring to teacher education*. Unpublished doctoral dissertation, Michigan State University.

Basturk, S., & Donmez, G. (2011).Examining pre-service teachers' pedagogical content knowledge with regard to curriculum knowledge. *International Online Journal of Educational Sciences*, 3 (2), 743-775.

Battista, M. T., & Clements, D. H. (1996). Students' understanding of three-dimensional rectangular arrays of cubes. *Journal for Research in Mathematics Education*. 27, 258-292.

Battista, M. T., Clements, D. H., Arnoff, J., Battista, K., & Borrow, C. V. A. (1998).Students' spatial structuring of 2D arrays of squares. *Journal for Research in Mathematics Education 29* (5), 503-532.

Battista, M. T. (1999). Fifth graders' enumeration of cubes in 3D arrays: Conceptual progress in an inquiry-based classroom. *Journal for Research in Mathematics Education*, *30*, 417 (-448.

Battista, M. T. (2002). Learning geometry in a dynamic computer environment; *Teaching Children Mathematics*,*8* (6), 333-339.

Battista, M.T. (2003).Understanding students' thinking about area and volume measurement.In Clements, D. H. (ed.), *2003 Yearbook*, *Learning and Teaching Measurement* (pp. 122-142).Reston, VA: NCTM.

Battista, M.T. (2004). Applying cognition-based assessment to elementary school students' development of understanding of area and volume measurement.*Mathematical Thinking and Learning, 6* (2), 185-204

Baturo, A., & Nason, R. (1996). Student teachers' subject matter knowledge within the domain of area measurement. *Educational Studies in Mathematics, 31*, 235-268

Begle, E. G. (1979).*Critical variables in mathematics education: Findings from a survey of empirical literature*. Washington, DC: Mathematical Association of America and National Council of Teachers of Mathematics.

Ben-Chaim, D., Lappan, G., & Houang, R. T. (1988). The effect of instruction on spatial visualization skills of middle school boys and girls. *American Educational Research Journal*, 25 (1), 51-71.

Benson, J. (1998), Developing a Strong Program of Construct Validation: A Test Anxiety Example. *Educational Measurement: Issues and Practice*, *17*, 10–17.

Binbaşıoğlu, C. (1995).*Türkiye'de Eğitim Bilimleri Tarihi*. Milli Eğitim Bakanlığı Yayınları. Ankara.

Blume, G. W., Galindo, E. & Walcott, C. (2007).*Performance in measurement andgeometry from the viewpoint of the Principles and Standards for School Mathematics*.In Kloosterman P. & F. K. Lester (Eds.), Results and interpretations of the 2003 Mathematics Assessment of the National Assessment of Educational Progress (pp.95– 138). Reston, VA: National Council of Teachers of Mathematics.

Bragg, P. & Outhred, L. (2000).Students' Knowledge of Length Units: Do They Know More than Rules about Rulers?Nakarahar, T. & Koyama, M: (Eds) In *Proceedings of the 24th Annual Conference of the International Group for the Psychology of Mathematics Education*, (Vol. 2, pp. 97-104). Hiroshima, Japan: PME.

Bode, R.K., & Wright, B.D. (1999). Rasch measurement in higher education. In J.C. Smart & W.G. Tierney (Eds.),*Higher education: Handbook of theory and research,* (Vol. XIV, pp. 287–316). New York: Agathon Press.

Borko, Hilda, Eisenhart, M., Brown, C. A., Underhill, R. G., & Agard, P. C. (1992). Learning to Teach Hard Mathematics : Do Novice Teachers and Their Instructors Give Up Too Learning to Teach Hard Mathematics : Do Novice Teachers and Their Instructors Give Up Too Easily ? *Journal for Research in Mathematics Education*, *23* (3), 194-222.

Borko, H., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education*, *23*, 194-222.

Boulton-Lewis, G. M., Wilss, L. A., & Mutch, S. L. (1996). An analysis of young children's strategies and use of devices for length measurement. *Journal of Mathematical Behaviour, 15*, 329–347.

Bruner, S. J. (1977).*The process of education.* Massachusetts: Harvard University Press.

Carpenter, T. P. (1975) Measurement concepts of first and second grade students.*Journal for Research in Mathematics Education*, *6*, 3-13

Carpenter, T. P, & Lewis, R. (1976) The development of the concept of a standard unit of measure in young children.*Journal for Research in Mathematics Education, 7*, 53-58

Carpenter, T. P., Corbitt, M. K., Kepner, H. S., Jr., Lindquist, M. M., & Reyes, R. E. (1981).*Results From the Second Mathematics Assessment of the National Assessment of Educational Progress*. Reston, VA: NCTM.

Carpenter, T. P., Fennema, E., Peterson, P. L., & Carey, D. A. (1988). Teachers' pedagogical content knowledge of students' problem solving in elementary mathematics. *Journal for Research in Mathematics Education, 19* (5), 385 (-401.

Carter, K. (1990). Teachers' knowledge and learning to teach. In W. R. Houston (Ed.),*Handbook of Research on Teacher Education* (pp. 291-310). New York: MacMillan.

Chappell, M. F. & Thompson, D. R. (1999). Perimeter or area?: Which measure is it? *MathematicsTeaching in the Middle School, 5* (1), 20-23

Clark, L.A., Watson, D., 1995. Constructing validity: Basic issues in scale development.*Psychological Assessment7* (3), 309-319.

Clements, D. H. & Stephan, M. (2004). Measurement in pre-K to grade 2 mathematics. In D. H. Clements, Sarama, J. & A.M. DiBiase (Eds.),*Engaging Young Children in Mathematics*, (pp. 299 – 317), Mahwah, NJ: Lawrence Erlbaum.

Clements, D. H., Battista, M. T., Sarama, J., Swaminathan, S., & McMillen, S. (1997). Students' development of length measurement concepts in a Logo-based unit on geometric paths. *Journal for Research in Mathematics Education*, 28 (1), 70-95.

Creswell, J. W. (2003).*Research Design: Quantitative, Qualitative, and Mixed Methods Approaches*. SAGE. Thousand Oaks. USA.

Crocker, L. & Algina, J. (1986).*Introduction to Classical and Modern Test Theory*. New York: CBS College Publishing.

Cooney, T.J. (1994).Research and teacher education: In search of common ground.*Journal for Research in Mathematics Education, 25* (6), 608-636.

Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna (Eds.),*Handbook of Test Development* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum Associates.

Ebel, R. L., & Frisbie, D. A. (1986).*Essentials of educational measurement* (4^{th}ed). Englewood Cliffs, NJ: Prentice-Hall

Eisenhart, M., Borko, H., Underhill, R., Brown, C., & Jones, D. (1993).Conceptual knowledge falls through the cracks : complexities of learning to teach mathematics for understanding. *Journal for Research in Mathematics Education, 24* (1), 8 (-40.

Ellis, S.,&Siegler, R.S. (1995, April) *Developmental changes in children's understanding of principles and procedures of measurement*. Paper presented at the biennial meeting of the Society for Research in Child Development, Indianapolis, IN.

Enochs, L. G., Gabel, D. L. (1984). Preservice elementary teachers' conceptions of volume. *School Science and Mathematics, 84* (8), 670-680

Even, R. (1993). Subject-Matter Knowledge and Pedagogical Content Knowledge: Prospective Secondary Teachers and the Function Concept. *Journal for Research in Mathematics Education, 24* (2), 94-116.

Fennema, E., & Franke, M. L. (1992). Teachers' knowledge and its impact. In D. A. Grouws (Eds.),*Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 147-164). New York: Macmillan.

Fenstermacher, G. D. (1994). The knower and the known: The nature of knowledge in research on teaching. In L. Darling-Hammond (Ed.),*Review of Research in Education* (Vol. 20, pp. 3–56). Washington, DC: AERA.

Fuller, R. A. (1997). Elementary teachers' pedagogical content knowledge of mathematics. *Mid-Western Educational Researcher*, 10 (2), 9-16.

Fuys, D., Geddes, D., & Tischler, R. (1988).*The van Hiele model of thinking in geometry among adolescents*. Reston, VA: National Council of Teachers of Mathematics.

Gess-Newsome, J. (1999). Pedagogical Content Knowledge: An Introduction and Orientation. In J. Gess-Newsome & N. G. Lederman (Eds.),*Examining Pedagogical Content Knowledge: The Construct and its Implications for Science Education* (pp. 3-17). Dordrecht, The Netherlands: Kluwer Publishing.

Goulding, M., Rowland, T. I. M., & Barber, P. (2002). Does it Matter ? Primary Teacher Trainees' Subject Knowledge in Mathematics. *British Educational Research Journal, 28* (5), 689-704.

Grimmett, P., & Mackinnon, A. (1992). Craft knowledge and the education of teachers. In G. Grant (Ed.),*Review of Research in Education*: (Vol.18 pp. 385 456). Washington, DC: American Educational Research Association.

Grossman, P. L. (1990).*The Making of a Teacher*. New York: Teachers College Press.

Grossman, P. L., Wilson, S. M., & Shulman, L. (1989). Teachers of Substance: Subject Matter Knowledge for Teaching. In M. C. Reynolds (Ed.),*Knowledge Base for the Beginning Teacher* (pp. 23-36). Oxford: Pergamon Press.

Grossman, P. (1991).*Mapping the terrain: Knowledge growth in teaching*. In H. C.Waxman & H. J. Walberg (Eds.), Effective Teaching: Current-Research (pp. 18-203). Berkley, CA: McCutchan Publishing

Gruijter, D. N. M. & van der Kamp, L. J. T. (2008).*Statistical Test Theory for the Behavioral Sciences*. Boca Raton, FL: Chapman & Hall / CRC

Haladyna, T. M. (1999).*Developing and validating multiple-choice test items.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Haladyna, T. M. (2004).*Developing and Validating Multiple- Choice Test Items* (3rd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, *12* (3), 3847.

Hambleton, R.K., & Swaminathan, H. (1984).*Item Response Theory: Principles and Applications.* Boston: Kluwer-Nijhoff Publishing.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991).*Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications.

Hershkowitz, R., & Vinner, S. (1984). Children's concepts in elementary geometry: A reflection of teacher's concepts? In Southwell, B. (Ed.),*Proceedings of the 8th Annual Conference of the International Group for th Psychology of Mathematics Education*, (Vol.1 pp. 63–69) Sydney, Australia: PME.

Hiebert, J. (1981). Cognitive development and learning linear measurement. *Journal for Research in Mathematics Education*, *12* (3), 197–211.

Hiebert, J. (1984). Why do some children have trouble learning measurement concepts? *Arithmetic Teacher*, *31* (7), 19-24.

Hill, H. C., Sleep, L., Lewis, J. M., & Ball, D. L. (2007). Assessing teachers' mathematical knowledge: What knowledge matters and what evidence counts. In F. K. Lester (Ed.),*Second Handbook of Research on Mathematics Teaching and Learning* (pp. 111-156). Charlotte, NC: Information Age Publishing.

Hill, H.C., Rowan, B., & Ball, D. L. (2005). Effects of Teachers ' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal*, *42* (2), 371 (-406.

Hill, H.C, Ball, D. L.,& Schilling, S. G. (2008). Unpacking Pedagogical Content Knowledge: Conceptualizing and Measuring Teachers' Topic-Specific Knowledge of Students. *Journal for Research in Mathematics Education*, *39* (4), 372 (-400.

Hirstein, J.J., Lamb, C.E.,& Osborne, A. (1978). Student misconceptions about area measure. *Arithmetic Teacher*, *25* (6), 10–16

Kamii, C. (2006). Measurement of length: How can we teach it better? *Teaching Children Mathematics, 13* (3), 154-158.

Kamii, C.,& F.B. Clark. 1997. Measurement of length: The need for a better approach to teaching. *School Science and Mathematics*, *97*, 116-121.

Kenney, P.A., & Kouba, V.L. (1997). What do students know about measurement? In P.A. Kenney & E. Silver (Eds.),*Results From the Sixth Mathematics Assessment of the National Assessment of Educational Progress* (pp. 141-163). Reston, VA: National Council of Teachers of Mathematics

Kehoe, J. (1995). Basic item analysis for multiple-choice tests. *Practical Assessment, Research & Evaluation,4* (10). Retrieved August 9, 2012 from http://PAREonline.net/getvn.asp?v=4&n=10.

Kidman, G., & Cooper, T. (1996). Children's perceptual judgement of area. *Proceedings of the 19th Annual Conference of the Mathematics Education Research Group of Australasia (MERGA)* (pp. 330-336). Melbourne: MERGA.

Kidman, G. & Cooper, T. J. (1997). Area integration rules for grades 4, 6, 8 students. In E.Pehkonen (Ed.),*Proceedings of the 21st Annual Conference of the International Group for the Psychology of Mathematics Education* (pp. 132– 143). Lahti, Finland: PME.

Kinach, B. (2002). Understanding and learning-to-explain by representing mathematics, *Journal of Mathematics Teacher Education*, *5*, 153-186.

Kordaki, M. (2003). The effect of tools of a computer microworld on students' strategies regarding the concept of conservation of area. *Educational Studies in Mathematics, 52*, 177-209,

Lee, E., & Luft, J. (2008). Experienced Secondary Science Teachers' Representation of Pedagogical Content Knowledge. *International Journal of Science Education, 30* (10), 1343-1363.

Lehrer,R.,Jenkins,M., &Osana,H. (1998).Longitudinal study of children's reasoning about space and geometry. InR.Lehrer& D. Chazan (Eds.),*Designing Learning Environments for Developing Understanding of Geometry and Space* (pp. 137 – 167). Mahwah, NJ: Erlbaum.

Lehrer, R., Jacobson, C., Kemeny, V., & Strom, D. A. (1999). Building upon children's intuitions to develop mathematical understanding of space. In E. Fennema & T. R. Romberg (Eds.),*Mathematics classrooms that promote understanding* (pp. 63-87). Mahwah, NJ: Erlbaum.

Lehrer,R.,Jacobson,C.,Thoyre,G.,Kemeny,V.,Strom,D.,Horvath,J.,Gance,S., &Koehler,M. (1998). Developing understanding of geometry and space in the primary grades. In R. Lehrer & D. Chazan (Eds.),*Designing Learning Environments for Developing Understanding of Geometry and Space* (pp. 169–200).Mahwah, NJ: Erlbaum.

Lehrer, R., Jaslow, L., & Curtis, C. (2003). Developing understanding of measurement in the elementary grades. In D. H. Clements & G. Bright (Eds.),*Learning and Teaching Measurement. 2003 Yearbook*. (pp. 100-121). Reston, VA: NCTM.

Lehrer, R. (2003). Developing understanding of measurement. In J. Kilpatrick, W. G. Martin& D. Schifter (Eds.),*A research companion to Principles and Standards for School Mathematics*, (pp. 179– 192).Reston, VA: NCTM.

Leinhardt, G. (1990). Capturing craft knowledge in teaching. *Educational Research, 19* (2), 18-25.

Leinhardt, G., & Smith, D. (1985). Expertise in mathematics instruction: Subject matter knowledge. *Journal of Educational Psychology*, *77* (3), 247–271.

Leinhardt, G. (1989). Math Lessons: A Contrast of Novice and Expert Competence. *Journal for Research in Mathematics Education, 20* (1), 52-75.

Linacre, J. M., & Wright, B. D. (2006).*WINSTEPS Rasch-model computer program* (Version 3.47) [computer software]. Chicago: MESA Press.

Lindquist,M. M., & Kouba,V. L. (1989). Measurement. In M. M. Lindquist (Ed),*Results from the Fourth Mathematics Assessment of the National Assessment of Educational Progress* (pp 35 (-43). Reston, VA: NCTM

Livingston, C., & Borko, H. (1990). High school mathematics review lesson: Expert-novice distinctions. *Journal for Research in Mathematics Education, 21* (5), 372-387.

Loevinger, J. (1957).Objective tests as instruments of psychological theory. *Psychological Reports. 3*, 635-694.

Loughran, J., Berry, A., & Mulhall, P. (2007). Pedagogical Content Knowledge: What Does It Mean to Science Teachers? In R. Pinto & D. Couso (Eds.),*Contributions From Science Education Research* (Vol. 3, pp. 93-105). Dordrecht: Springer.

Loughran, J., Mulhall, P., & Berry, A. (2008). Exploring Pedagogical Content Knowledge in Science Teacher Education. *International Journal of Science Education*, *30* (10), 1301-1320.

Ma, L. (1999).*Knowing and Teaching Elementary Mathematics: Teachers' Understanding of Fundamental Mathematics in China and The United States*. Mahwah, NJ: Lawrence Erlbaum Associates.

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.),*Examining pedagogical content knowledge: The construct and its implica-tions for science education (*pp. 95-132). Dordrecht, The Netherlands: Kluwer Academic.

Maher, C.A. and Beattys, C. B. (1986).Examining the Construction of area and its Measurement by Ten to Fourteen Year old Students. E. Lansing, G. Lappan, and R. Even (Eds),*In Proceedings of the 8th Annual Conference of the International Group for the Psychology of Mathematics Education*, (Vol.1, pp. 163–168) Sydney, Australia: PME.

Mapolelo, D. C. (1999). Do pre-service primary teachers who excel in mathematics become good mathematics teachers? *Teaching and Teacher Education*, *15* (6), 715-725.

Martin, M.O., Mullis, I.V.S., & Foy, P. (2008).*TIMSS 2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades.* TIMSS & PIRLS International Study Center, Chestnut Hill, MA. Boston College.

Martin, W. G., & Strutchens, M. E. (2000). Geometry and measurement. In E. A. Silver & P. A. Kenney (eds.),*Results from the Seventh Mathematics Assessment of the National Assessment of Educational Progress*. Reston, VA: National Council of Teachers of Mathematics.

Mason, J., & Spence, M. (1999). Beyond mere knowledge of mathematics: The importance of knowing-to act in the moment. *Educational Studies in Mathematics, 38*, 135-161.

Mayberry, J. W. (1983). The van Hiele levels of geometric thought in undergraduate preservice teachers. *Journal for Research in Mathematics Education, 14* (1), 58-69.

Messick, S. (1989). Validity. In R. Linn (Ed.),*Educational measurement* (3rd ed., pp. 13–103). Washington, DC: American Council on Education/Macmillan.

Mitchell, J., & Williams, S. E. (1993).*Expert: Novice Difference in Teaching with Technology.* Distributed by ERIC Clearinghouse

Mitchelmore, M. C. (1983).Children's learning of geometry: Report ofa cooperative research project. *Caribbean Journal of Education, 10,* 179-228.

Monk, D. H. (1994). Subject matter preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, *13* (2), 125-145.

Monk, D. H. & King, J. A. (1994). Multi-level teacher resource effects in pupil performance in secondary mathematics and science: The case of teacher subject matter preparation. In R. G. Ehrenberg (Ed.),*Choices and consequences: Contemporary policy issues in education*, (pp. 29-58). Ithaca, NY: ILR Press.

Mullis, I., Dossey, J., Owen, E., & Phillips, G. (1991).*The state of mathematics achievement: NAEP's 1990 assessment of the nation and the trial assessment of the states (NCES Publication No. 91-1259).* Washington, DC: National Center for Education Statistics.

National Council of Teachers of Mathematics [NCTM]. (2000).*Principles and standards for school mathematics*. Reston, VA: NCTM.

Nitabach, E. and Lehrer, R. (1996). Developing spatial sense through area measurement. *Teaching Children Mathematics, 2,* 473 (-476.

Nunes, T., Schliemann, A. & Carraher, D. (1993).*Street Mathematics and School Mathematics*. Cambridge, Cambridge University Press

Nunes, T., Light, P., and Mason, J.H. (1993). Tools for thought: The measurement of length and area. *Learning and Instruction, 3*, 39-54.

Nunnally, J. C. (1967).*Psychometric theory*. New York McGraw-Hill

Outhred, L. & Mitchelmore, M. (1996). Children's intuitive understanding of area measurement. In L. Puig& A. Gutierrez (Eds.*) Proceedings ofthe 20thConference ofthe international groupfor the Psychology of Mathematics Education*, (Vol. 4, pp. 91-98). Valencia, Spain: PME

Outhred, L. (1993). *The development in young students of concepts of rectangular area measurement.* Unpublished PhD dissertation, Macquarie University, Australia.

Outhred, L., & Mitchelmore, M. C. (2000). Young children's intuitive understanding of rectangular area measurement. *Journal for Research in Mathematics Education, 31*, 144-167.

Park, S., Oliver, J.S. (2007). Revisiting the conceptualization of pedagogical content knowledge (PCK) :PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education,38*, 261-284.

Piaget, J., Inhelder, B., & Szeminska, A. (1981).*The child's conception of geometry*. New York: Norton.

Raghavan, K., Sartoris, M. L., & Glaser, R. (1998). Interconnecting science and mathematics concepts: Area and volume. In R. Lehrer & D. Chazan (Eds.), *Designing Learning Environments for Developing Understanding of Geometry and Space* (pp. 267-295). Hillsdale, NJ: Erlbaum.

Reece, C. S., & Kamii, C. (2001). The measurement of volume: Why do young children measure inaccurately? *School Science and Mathematics, 101*, 356-361.

Reynolds, A., & Wheatley, G. H. (1996). Elementary students' construction and coordination of units in an area setting. *Journal for Research in Mathematics Education, 27* (5), 564-581.

Ryan, J. & Williams, J. (2007).*Children's Mathematics 4-15: Learning from Errors and Misconceptions*. Open University Press

Saiz, M. (2003).Primary teachers' conceptions about the concept of volume: The case of volume measurable objects. In N. A. Pateman, B. J. Dougherty, and J. T. Zilliox (Eds),*Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education*, (Vol. 4, pp. 95 –102). Honolulu, Hawaii: PME

Sarama, J., Clements, D. H., Swaminathan, S., McMillen, S., & Gonzalez Gomez, R. M. (2003). Development of mathematical concepts of two-dimensional space in grid environments: An exploratory study. *Cognition and Instruction, 21* (3), 285-324.

Schilling, S.G. & Hill, H.C. (2007). Assessing Measures of Mathematical Knowledge for Teaching: A Validity Argument Approach. Measurement: *Interdisciplinary Research and Perspectives (5),* 2 (-3, 70-80.

Segall, A. (2004) Revisiting pedagogical content knowledge: the pedagogy of content/the content of pedagogy. *Teaching and Teacher Education*, *20*, 489–504 .

Senk, S., & Thompson, D. (2003).*Standards-based school mathametics curricula: What are they? What do students learn?* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Shulman, L. (1986). Those Who Understand : Knowledge Growth in Teaching. *Educational Researcher, 15* (2), 4-14.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57* (1), 1-22.

Simon, M. A. & Blume, G. (1994). Building and understanding multiplicative relationships: A study of prospective elementary teachers. *Journal for Research in Mathematics Education, 25,* 472 (-494.

Simon, M. A. (1995). Investigating the development of multiplicative reasoning. *Journal for Research in Mathematics Education, 26*, 282-287.

Skemp, R. (1986).*The psychology of learning mathematics*. London: Penguin Books.

Smith, J., P., Tan-Sisman, G., Figueras, H., Lee, K., Dietiker, L., & Lehrer, R. (2008).*Assessing curricular contributions to poor measurement learning*. In Research symposium (90-min session) at the National Council of Teachers of Mathematics (NCTM) 2008 annual meeting and exposition: Becoming certain about uncertainty. Salt Lake City, Utah.

Smith, D. (1999). Changing our teaching: The role of pedagogical content knowledge elementary science. In J. Gess-Newsome & N. Lederman (Eds.),*Examining Pedagogical Content Knowledge*. London, UK: Kluwer Academic Publishers.

Stephan, M., & Clements, D.H. (2003). Linear and area measurement in prekindergarten to grade 2. In D. H. Clements & G. Bright (Eds.),*Learning and teaching measurement. NCTM 2003 Yearbook* (pp. 3–16). Reston, VA: NCTM.

Strutchens, M. E., Martin, W. G., & Kenney, P. A. (2003). What students know about measurement: Perspectives from the NAEP. In D. H. Clements & G. Bright (Eds.),*Learning and Teaching Measurement: NCTM 2003 Yearbook* (pp. 197-208). Reston: NCTM.

Stylianides, A. J., & Ball, D. L. (2008). Understanding and describing mathematical knowledge for teaching: Knowledge about proof for engaging students in the activity of proving. *Journal of Mathematics Teacher Education, 11,* 307-332.

Szilagyi,J. (2007).Youngchildren'sunderstandingsoflengthmeasurement:Adevelopmental progression. Ph.D. dissertation, State University of New York at Buffalo, NY.

Şişman, M., Acat, M. B., Aypay, A. & Karadağ, E. (2011). TIMSS 2007 Ulusal Fen Raporu: 8. Sınıflar. Ankara: EARGED Yayınları.

Tierney, C., Boyd, C. & Davis, G., (1990). Prospective primary teachers' conceptions of area. In G. Booker, P.Cobb, and T.N. de Mendicuti (Eds),*Proceedings of the 14th Conference of the International Group for the Psychology of Mathematics Education with the North American Chapter 12th PME-NA*, (Vol 2, pp. 307-315). Mexico: PME

Tirosh, D. (2000). Enhancing Prospective Teachers' Knowledge of Children's Conceptions : The Case of Division of Fractions. *Journal for Research in Mathematics Education, 31* (1), 5-25.

Vacc, N. N., & Bright, G. W. (1999). Elementary Preservice Teachers' Changing Beliefs and Instructional Use of Children ' s Mathematical Thinking. *Journal for Research in Mathematics Education, 30* (1), 89-110.

van de Walle, J. A. (2007).*Elementary And Middle School Mathematics: Teaching Developmentally* (6th ed.). Boston: Pearson Education, Inc.

van Driel, J.H., Verloop, N. & de Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching, 35* (6) 673-695

Vergnaud, G. (1983). Multiplicative structures. In R. Lesh and M. Landau (Eds.),*Acquisition of Mathematical Concepts and Processes* (pp. 127-174). Orlando: Academic Press.

Verloop, N., Driel, J. V., & Meijer, P. (2001). Teacher knowledge and the knowledge base of teaching. *International Journal of Educational Research, 35*, 441 (-461.

Walsh, B. W., & Betz, N. E. (1995).*Tests and assessment* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall

Wilson, P.,& Rowland, R. (1993). Teaching Measurement. In P. Jenner, (Ed.),*Research Ideas for the Classroom. Early Childhood Mathematics*, (pp. 171-191). New York: Macmillan.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, *14*, 97-116.

Zacharos, K. (2006). Prevailing educational practices for area measurement and students‟ failure in measuring areas. *Journal of Mathematical Behavior*,*25*, 224-239.

Zembat, İ. Ö. (2010) Ölçme Temel Bileşenleri ve Sık Karşılaşılan Kavram Yanılgıları. Bingölbali, E., & Özmantar, M.,F. (Eds.) *İlköğretimde karşılaşılan Matematiksel Zorluklar ve Çözüm Önerileri,* (pp.127-154). Ankara: Pegem Akademi

APPENDICES

APPENDIX A
INITIAL FORMS OF ITEMS

İletişim için:yesen@metu.edu.tr

Test_0

PEDAGOJİK ALAN BİLGİSİ TESTİ
(ÖLÇME)

ODTÜ

ADI :...

SOYADI :...

CİNSİYET : KADIN ☐ ERKEK: ☐

ÜNİVERSİTE :.............................BÖLÜM : ....

GENEL NOT ORTALAMASI : ....

DÖNEM : ....

GENEL AÇIKLAMA

1.Bu sınavda her katılımcıya bir soru kitapçığı verilecektir.

2.Bu soru kitapçığı her birine **5 seçenek sunulan 29 sorudan** oluşmaktadır.

3.Testin tümü için öngörülen cevaplama süresi **60 dakika'dır.**

4.Cevaplamaya istediğiniz sorudan başlayabilirsiniz. Cevaplarınızı kitapçık üzerine işaretlemeyi unutmayınız.

5.Açık uçlu olarak soruları mümkün olduğu kadar detaylı cevap vermeye çalışınız.

6.Bu ölçeğin sonuçları sadece araştırma amaçlı olup hiçbir şekilde notlarınızı **etkilemeyecektir.**

Teşekkürler.

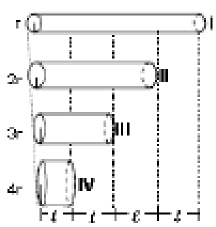
İLETİŞİM BİLGİLERİ
Yasemin ESEN
ODTÜ Eğitim Fakültesi
İlköğretim Matematik Eğitimi Bölümü
06531 Çankaya- ANKARA
e-mail: yaseminesen@gmail.com

Tel: 0 312 210 40 65

- 1 -

SORULAR

1)



İçi boş r,2r,3r ve 4r yarıçaplı silindirlerin boyları şekildeki gibidir. Silindirlerin her birisi yarıya kadar su ile doldurulduğu zaman aldıkları su miktarlarını sıralayınız.

(DPY 2005/9-10-11. Sınıf Soru 2)

Soruya ***"I < II < III < IV"*** gibi yanıt veren bir öğrenci silindirin hacmine etki eden değişkenlerden sadece birini dikkate alarak sıralama yapmıştır.

Öğrencinin yaşadığı bu güçlüğü aşması için aşağıda verilen etkinliklerden hangisi ya da hangileri kullanılabilir?

A) Yarıçapları aynı yarı dolu silindir ile tam dolu silindirin hacimlerini karşılaştırmak.
B) Yükseklikleri farklı taban alanları aynı silindirlerin hacimlerini karşılaştırmak
C) Taban alanları aynı yükseklikleri farklı kare prizmaların hacimlerini karşılaştırmak
D) Bir taban aynı aynı, yükseklikleri farklı kare prizma ile dikdörtgenler prizmasının hacimlerini karşılaştırmak
E) Yükseklikleri aynı taban alanları farklı kare prizmaların hacimlerini karşılaştırmak

Gerekçe:..
....

Soru ile ilgili öneri:...........................

....
....

2)

Aşağıda içlerine üçer tanetop tam olarak sığdırılabilen iki silindir verilmiştir. Bu kutulardaki güvenlik şeritlerinden birisi *Şekil I*'deki gibi dikine, diğeri ise *Şekil II*'deki gibi kapağın çevresine yerleştirilmiştir. Öğrencilerinden bu iki şerit arasında nasıl bir ilişki olabileceğini soran öğretmen aşağıdaki yanıtları almıştır.

Şekil I *Şekil II*

Aşağıda verilen öğrenci yanıtlarından hangisi doğrudur?

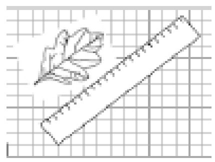
A) Kapağın çevresine sarılan şerit, dik şeritten daha uzundur.
B) Kapağın çevresine sarılan şerit, dik şeritten daha kısadır.
C) Kapağın çevresine sarılan şerit ile dik şeridin boyu eşittir.
D) Sayısal bir değer olmadığı için bir şey söylenemez.
E) Yeterince büyük toplar kullanıldığında Şekil II'deki şerit, Şekil I'deki şeritten daha uzun olur.

Gerekçe:..
....

Soru ile ilgili öneri:...........................

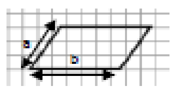
....
....

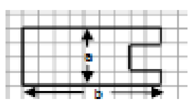
3)

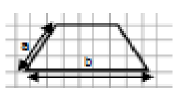


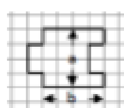
Bir öğrenci şekildeki gibi verilen bir yaprağın alanını eni ile boyunu çarparak bulabileceğini düşünmektedir.

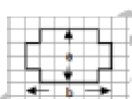
Seçeneklerde verilen ifadelerden hangisinde doğru sonuca ulaşması, öğrencinin yaptığı *"verilen uzunluklar çarpılarak alan bulunur"* şeklinde bir düşünce geliştirmesine yol açmış olabilir?

A)  Paralelkenarsal bölgenin alanını a ve b uzunluklarını çarparak bulma

B)  Verilen bölgenin alanını a ve b uzunluklarını çarparak bulma

C)  Yamuksal bölgenin alanını a ve b uzunluklarını çarparak bulma

D)  Verilen bölgenin alanını a ve b uzunluklarını çarparak bulma

E)  Verilen bölgenin alanını a ve b uzunluklarını çarparak bulma

Gerekçe:...

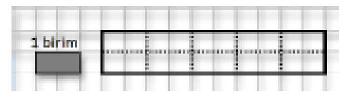
....

Soru ile ilgili öneri:..........................

....

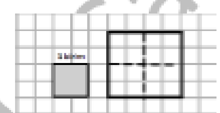
....

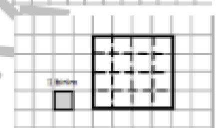
4)

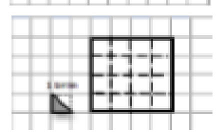


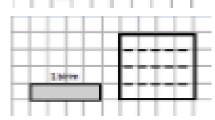
Alan kavramını işleyen bir öğretmen öğrencilerinden yukarıdaki dikdörtgensel bölgenin alanını yanında verilen birim cinsinden ölçmelerini ister

Öğrencinin yaşadığı bu zorluğu aşması için aşağıda verilen örneklerden hangisi ya da hangileri kullanılabilir?

I)  *Alan= 4 Birim*

II)  *Alan= 16 Birim*

III)  *Alan= 32 Birim*

IV)  *Alan= 4 Birim*

A) Hepsi B) I, II, III C) I, III, IV

D) II, IV E) Yalnız II

Gerekçe:...

....

Soru ile ilgili öneri:..........................

....

....

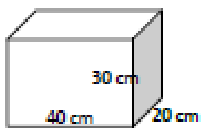
- 2 -

5) Sınıf:6

Öğrenme Alanı: Ölçme

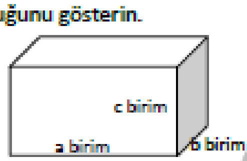
Kazanım: Dikdörtgenler prizması, kare prizma ve küpün yüzey alanlarını hesaplar.

I. Aşağıda verilmiş olan sürpriz kutunun bütün yüzleri ayna ile kaplanmak istenirse, kaç cm^2 aynaya ihtiyaç olur?



II. Boyutları 20 cm x 30 cm olan bir kartondan kesilerek prizma yapılacaktır. Yüzey alanı en fazla olacak şekilde oluşturulan prizmanın ayrıt uzunluklarını bulun.

III. Ayrıt uzunlukları yukarıdaki gibi verilen dikdörtgenler prizmasının yüzey alan formülünün " *2(a.b) +2(b.c) + 2 (a.c)* " olduğunu gösterin.



MEB Matematik Öğretim Programı dikkate alındığında yukarıdaverilen sorulardan hangi ya da hangileri belirtilen kazanımı 6. sınıf seviyesinde ölçmeye yönelik <u>değildir</u>?

A) Yalnız I
B) Yalnız II
C) Yalnız III
D) I, III
E) I, II, III

Gerekçe:..
....

,

Soru ile ilgili öneri:.........................

....
....

6)

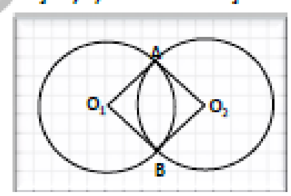
Çember ve daire konusunu işleyen bir öğretmen sınıfta majör ve minör yayları tanıttıktan sonra öğrencilerinden **merkezdeş iki çember çizmelerini; bu çemberlere ait minör yay uzunlukları ve yay ölçüleri** ile ilgili değerlendirme yapmalarını istemiştir.

Öğrenci I) Merkezdeş iki çemberde minör yaylardan biri diğerinden kısa ise, kısa olan minör yayın ölçüsü diğerinden küçüktür.

Öğrenci II) Merkezdeş iki çemberde minör yayların uzunlukları eşit ise, yay ölçüleri eşittir.

Öğrenci III) Şekilde belirtildiği gibi merkez açılar, aynı yayı gördükleri için yay uzunlukları eşittir.

Öğrencinin çizdiği şekil:



Öğrenci IV) Merkezdeş iki çemberde ölçüleri aynı merkez açıların gördükleri yayların ölçüleri aynıdır.

Yukarıda verilen öğrenci önermelerinden hangi ve ya hangileri <u>her zaman doğrudur</u>?

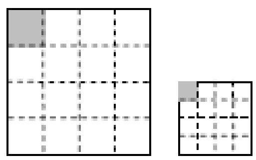
A) I, II, III, IV
B) I, II, IV
C) I, II
D) Yalnız I
E) Yalnız IV

Gerekçe:...
....

Soru ile ilgili öneri:.........................

....
....

7)



Şekil I Şekil II

Bir öğrenci, yukarıda verilen iki bölgenin taralı olarak belirtilmiş birimlerle ölçüldükleri zaman alanlarının aynı olacağını düşünmektedir.

Bu şekilde düşünen bir öğrenci ile aşağıdaki karşılaştırmaları yapan öğrencilerden hangisi ya da hangileri arasında benzer bir düşünce yapısı vardır?

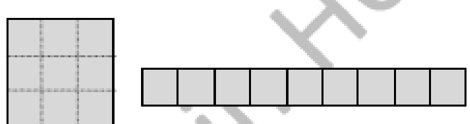
I) **Öğrenci I**: "Her ikisi de eşit uzunluktadır."

Şekil I Şekil II

II) **Öğrenci II**: "Şekil II Şekil I'den daha uzundur."

Şekil I Şekil II

III) **Öğrenci III**: "Şekil II Şekil I'den daha fazla alana sahiptir."



Şekil I Şekil II

A) Yalnız I
B) Yalnız II
C) Yalnız III
D) I ve II
E) I ve III

Gerekçe:..
....

Soru ile ilgili öneri:.........................
....
....

8)

Ölçme kavramını işleyen bir öğretmen dersine başlamadan önce öğrencilerine resimdeki gibi bir kova gösterir.

Öğretmenin aşağıdakiifadelerinden hangisi kovanın bir özelliğine ilişkin ölçüm değildir?

A) Kovanın simetri eksenince uzatılan ipin kovanın içinde kalan kısmının uzunluğunu ölçmek.

B) Kovanın ağzında madeni para yuvarlayıp paranın kaç tur yaptığını saymak.

C) Sıvı içinde saniyede 2 metre çöken bir aracın, sıvı dolu kova içine attıktan sonra tabana çökene kadar geçen süreyi hesaplamak.

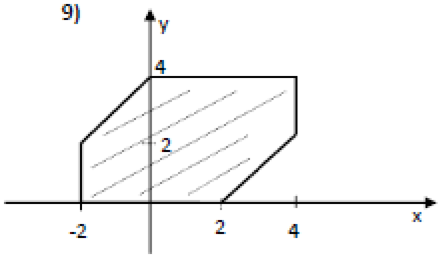
D) Kovanın içini doldurduktan sonra esnek bir iple asıp ipteki esneme miktarını ölçmek.

E) Kovayı yarım litrelik sularla doldurup kaç pet şişe boşaltıldığını saymak.

Gerekçe:..
....

Soru ile ilgili öneri:.........................
....
....

9)



Yukarıda verilen şeklin x-ekseni etrafında 180^0 döndürülmesi ile elde edilen şeklin hacmi kaç birim^3'tür?

A) $\frac{80\pi}{3}$ B) $\frac{108\pi}{3}$ C) $\frac{160\pi}{3}$
D) $\frac{216\pi}{3}$ E) $\frac{240\pi}{3}$

Gerekçe:..
....

Soru ile ilgili öneri:.........................

....
....

10)

Koninin yüzey alan bağıntısını öğrencilerle birlikte sınıfta elde edecek bir öğretmen, konuya başlamadan önce öğrencilerindeaşağıda verilen kavramlardan hangi ya da hangilerinde bilgi eksikliği olmadığından emin olmalıdır?

I. Dairenin alanı
II. Daire dilimin alanı
III. Çemberin çevresi
IV. Yükseklik

A) I, II, III, IV B) I, II, III C) I, II
D) III, IVE) Yalnız II

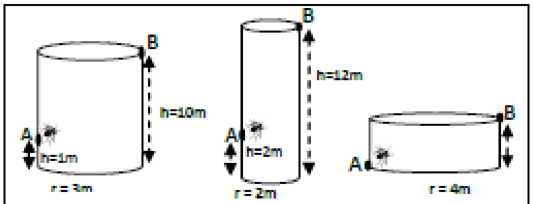
Gerekçe:..
....

Soru ile ilgili öneri:.........................

....
....

İletişim için:yesen@metu.edu.tr

11)



Şekil I *Şekil II* *Şekil III*

Bir karınca, yükseklikleri ve yarıçapları verilmiş silindirler üzerinde belirtilen A noktalarından B noktalarına silindir yüzeyinden yürüyerek gidecektir.

Karıncanın yürüyeceği uzaklıkları büyükten küçüğe doğru sıralayınız.

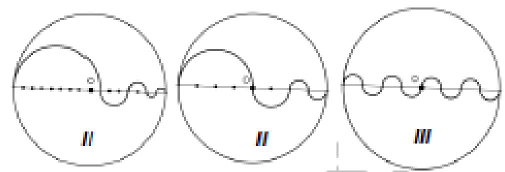
A) I, II, III **B)** III, II, I **C)** II, I, III
D) I, III, II **E)** III, I, II

Gerekçe:...
....

Soru ile ilgili öneri:...........................
....
....

12)

Aşağıda verilen özdeş çemberlerin içine çizilen eğrilerin uzunlukları arasında nasıl bir ilişki vardır?

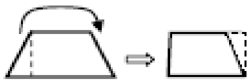


A) I > II > III **B)** I > II=III **C)** I = II> III
D) II= III > I **E)** I= II =III

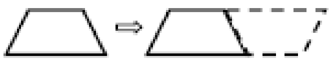
Gerekçe:...
....

Soru ile ilgili öneri:...........................
....
....

13) Yamuksal bölgenin alanını işlemekte olan bir öğretmen, öğrencilerinden yamuksal bölgenin alanını bulmaya yönelik farklı stratejiler geliştirmelerini istemiştir.

I) 

Kesik çizgili yerden kesip, şekildeki gibi birleştirerek dikdörtgensel bölgenin alan formülünü kullanmak.

II) 

Şeklin eşini şekildeki gibi ekleyerek paralelkenarsal bölgenin alan formülünü kullanıp, yarısını almak.

III) 

Şekilde verildiği gibi yamuğu iki üçgensel bölgeye ayırıp, üçgensel bölgenin alan formülünü kullanıp, alanlarını toplamak.

IV) 

Kesik çizgili yerden kesip, şekildeki gibi birleştirerek paralelkenarsal bölgenin alan formülünü kullanmak.

V) 

Kesik çizgili yerden kesip, şekildeki gibi birleştirerek dikdörtgensel bölgenin alan formülünü kullanmak.

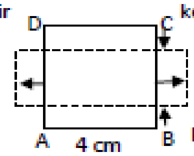
Öğrencilerden gelen yukarıdaki yöntemlerden hangi ya da hangileri herhangi bir yamuksal bölgenin alanını bulmaya yönelik olarak genellenebilir?

A) I, II, III, IV, V B) II, III, IV C) I, III, V D) II, III E) Yalnız III

Gerekçe:...
....

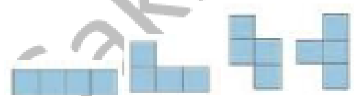
Soru ile ilgili öneri:.......................
....
....

14) Sınıfta alan konusu işlenirken bir öğrenci, bir geometrik şeklin çevresinin artması durumunda şeklin kapladığı alanın artacağını söylemiştir.

I) Bir 

kenar uzunluğu 4 cmolan bir karenin karşılıklı kenar uzunlukları 1 cm uzayıp diğer kenarların 1 cm kısalması durumunda alan azalır.

II) Aşağıda verilen 4 karelilerde farklı şekiller oluşturmasına rağmen alanları sabit kalmaktadır.



III) Sabit uzunlukta bir iple uçları bağlandıktan sonra oluşturulan kapalı geometrik şekillerin çevreledikleri bölgelerin alanlarıfarklıdır.

IV) 5 birim kare ile oluşturulabilecek şekillerin kapladıkları alanlar sabit kalmaktadır.

Öğrencinin düşüncesine karşıt örnek oluşturmak için yukarıda verilen durumlardan hangisi ya da hangileri kullanılabilir?

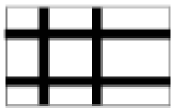
A) I, II, III, IV B) II, IV C) I, III D) Yalnız III E) Yalnız IV

Gerekçe:...
....

Soru ile ilgili öneri:.......................
....
....

15)

Alan konusunu işleyen bir öğretmen öğrencilerine aşağıdaki soruyu soruyor.



Kenar uzunlukları 200 m ve 100 m olan dikdörtgen şeklindeki bir arsa, kenarlarına dik olan 2 m genişliğinde yollar açılarak şekildeki gibi 9 parsele ayrılmıştır.Açılan yolların toplam alanı kaç m^2 dir? *(DPY/ 2005 – 7. Sınıf Soru 21)*

Soruya 1200 m^2 yanıtını veren öğrencinin, aynı arsanın, aynı genişlikteki 3 yatay ve 3 dikey yol ile 16 parsele ayrılması durumunda yolların toplam alanını kaç m^2bulması beklenir?

A) 1796 B) 1791 C) 1788 D) 1764 E) 1800

Gerekçe:...

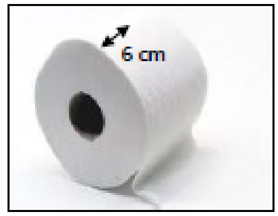
....

Soru ile ilgili öneri:...

....

....

16)



Yanda verilen tuvalet kâğıdının sarıldığı rulonun yarıçapı 2 cm ve sarılan tuvalet kâğıdının kalınlığı 6 cm' dir.

Kâğıdın ruloya 200 kez dolandığı düşünülürse, tuvalet kâğıdının uzunluğu <u>yaklaşık</u> kaç metredir?

A) 24 B)36 C)60 D) 72 E)96

Gerekçe:...

....

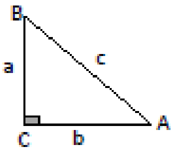
Soru ile ilgili öneri:...

....

....

17)

Pisagor bağıntısı:



Yandaki gibi verilen bir dik üçgende dik kenarların kareleri toplamı en uzun kenarın (hipotenüsün) karesine eşittir.
$c^2=a^2+b^2$

MEB 8. Sınıf Matematik Öğretim Programını dikkate alan bir öğretmen, sınıf ortamında aşağıdaki kavramlardan hangi ya da hangilerini kullanarak Pisagor Bağıntısını elde edebilir?

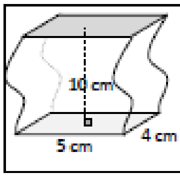
I) Karesel bölgenin alanı
II) Dairenin alanı
III) Düzgün altıgensel bölgenin alanı
IV) İkizkenar üçgensel bölgenin alanı

A) Yalnız I B) I, II C) I, IV
D) I, II, III E) Hepsi

Gerekçe:..
....

Soru ile ilgili öneri:..
....
....

18)



Yandaki şekilde yüksekliği 10 cm, taban ayrıtları 4cm ve 5cm olan 3 boyutlu bir cisim verilmiştir.

Verilen 3 boyutlu cismin hacmi ile ilgili olarak aşağıda verilen ifadelerden hangisi doğrudur?

A) Cismin yan yüzleri birbirine paralel ise, cisim prizmadır ve (*5 cm x 4 cm x 10 cm*) işlemi ile hacmini 200 cm^3 olarak bulabiliriz.

B) Cismin yan yüzleri birbirine paralel ise, cisim prizmadır fakat yan yüzler eğri olduğu için (*5 cm x 4 cm x 10 cm*) işlemini kullanamayız.

C) (*5 cm x 4 cm x 10 cm*) işlemini kullanabilmek için yan yüzlerinin tabana dik olması gerekir. Yan yüzleri dikleştirmek için yukarı doğru çekerek uzattığımızda cismin yüksekliği 10 cm'den uzun olacağından hacmi 200cm^3'ten fazladır.

D) Cismin yan yüzleri birbirine paralel olsa bile, cisim prizma değildir fakat (*5 cm x 4 cm x 10 cm*) işlemi ile hacmini 200cm^3 olarak bulabiliriz.

E) Cismin yan yüzleri birbirine paralel olsa bile, cisim prizma değildir. Bu sebeple (*5 cm x 4 cm x 10 cm*) işlemini kullanarak hacim bulamayız.

Gerekçe:..
....

Soru ile ilgili öneri:..
....
....

19)

Sınıfta alan konusunu işleyen bir öğretmen, öğrencilerinden aşağıda farklı büyüklükteki birimler cinsinden değerleri verilen geometrik şekillerin alanlarını karşılaştırmalarını istiyor.

	Şekil I	Şekil II
Değer:	25 birim kare	30 birim üçgen

Öğrenci I: Şekil II, Şekil I'den büyüktür.

	Şekil I	Şekil II
Değer:	26 birim kare	15 birim kare

Öğrenci II: Şekil I, Şekil II'den büyüktür.

	Şekil I	Şekil II
Değer:	12 cm^2	9 cm^2

Öğrenci III: Şekil I, Şekil II'den büyüktür.

Yukarıdakisayısal bilgilere dayanarak yapılan öğrenci önermelerinden hangisi ya da hangileri her zaman doğrudur?

A) Öğrenci I, Öğrenci II, Öğrenci III
B) Öğrenci II, Öğrenci III
C) Yalnız Öğrenci I
D) Yalnız Öğrenci II
E) Yalnız Öğrenci III

Gerekçe:...
....

Soru ile ilgili öneri:...
....
....

20)

Sınıfta tahmin etme stratejileri üzerine etkinlik yapan bir öğretmen, öğrencilerine farklı durumlar sunmuş ve öğrencilerinden bu durumlarla ilgili tahmin yapmalarını istemiştir. Sınıftan gelen tahminlerden 5 tanesi aşağıda verilmiştir.

Hangi öğrenci diğerlerinden farklıtahmin stratejisi kullanmaktadır?

A) *Bir serayı kapatmak için gereken şeffaf malzeme için*: Seranın 1 metre eni için 147 m^2 şeffaf gereklidir, seranın eni yaklaşık 60 metredir. O zaman yaklaşık 60x150 m^2 şeffaf gereklidir.

B) *Bir kavanozdaki pirinç sayısı için:* Bir kaşık pirinçte 120 pirinçtanesi, bir kavanozda yaklaşık 50 kaşık pirinç vardır. O zaman yaklaşık 50x120 pirinç tanesi vardır.

C) *Bir otele gereken fayans miktarı için:*Bir oda için yaklaşık 10 kutu fayans gereklidir, her katta 15 oda vardır. O zaman 5 katlı bir otel için yaklaşık 10x75 kutu fayans gereklidir.

D) *İki nokta arasındaki mesafe için:* Bir dakikada 50 metre yürüyebiliyorum, yolu yaklaşık 15 dk'da yürüyebilirim. O zaman iki nokta arasındaki mesafe yaklaşık 15x50 m'dir.

E) *Bir depoda bulunan toplam su miktarı için:* Deponun her 50 cm yüksekliğinde yaklaşık 5 ton su bulunmaktadır. Deponun yüksekliği yaklaşık 6 metredir. O zaman depoda yaklaşık 12x5 ton su bulunmaktadır.

Gerekçe:...
....

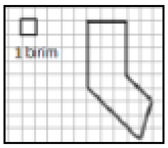
Soru ile ilgili öneri:...
....
....

İletişim için:yesen@metu.edu.tr

21) MEB 6Matematik Öğretim Programı dikkate alındığında, alan konusu ile ilgili aşağıdaki verilen sorulardan hangileri 6. Sınıf seviyesine uygundur?

I) Verilen dönüşümü yapın. 810 dekar= .... ar

II)

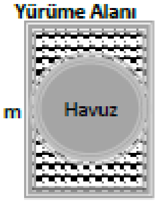


Yandaki şeklin alanı kaç br^2'dir.

III)

Yürüme Alanı

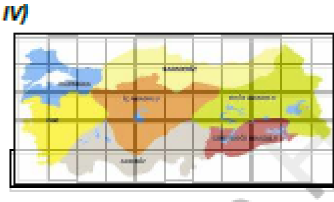
400 m



200 m

Yanda verilen kare şeklindeki bir bölgeye park ve yürüme alanı yapıldıktan sonra geriye kalan alan kaç m^2 olur?

IV)



Türkiye'nin toplam yüzölçümü 814.578 km^2 iseAkdeniz Bölgesi'nin yüz ölçümü yaklaşık kaç dekar'dır?

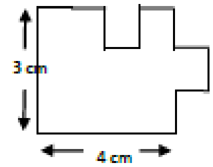
A) I, II, III, IV B) I, II, III C)I, II, IV
D) II, III, IV E) II, III

Gerekçe:...
....

Soru ile ilgili öneri:..
....
....

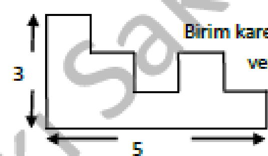
22)

Öğretmen: Sonucu kaç buldun?
Öğrenci: 12 cm^2
Öğretmen: Nasıl buldun?
Öğrenci: Verilen kenar uzunluklarını çarptım.
Öğretmen: Sence yöntemin doğru mu?
Öğrenci: (Düşünür).Ama birim karelerle kaplasam da aynı sonuç çıkıyor.

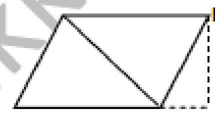


Verilen şeklin alanı ile ilgili olarak öğrencisi ile yukarıdaki gibi bir diyalog geçen öğretmenin aşağıdakilerden hangisini yaptırması öğrencide oluşan bu algının giderilmesine yardımcı olur?

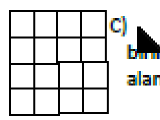
A)



Birim karelerikullanarak verilen şeklin alanını buldurmak

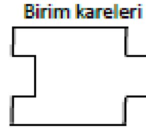


B) Şeklin kenar uzunluklarını cetvelle ölçtürdükten sonra şeklin alanını buldurmak

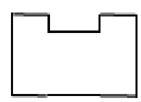


C) Üçgen şeklindeki birimler kullanarak şeklin alanını buldurmak

D) Birim kareleri kullanarak verilen şeklin alanını buldurmak



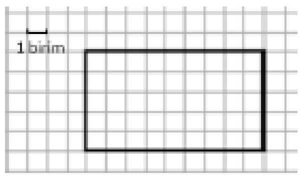
E) Şeklin kenar uzunluklarını cetvelle ölçtürdükten sonra şeklin alanını buldurmak



Gerekçe:...
....

Soru ile ilgili öneri:..
....
...

23)



$$\frac{9\ \text{birimkare}}{1\ \text{satırda}} = \frac{x\ \text{birimkare}}{5\ \text{satırda}} \rightarrow \frac{9x5}{1x5} = \frac{x}{5}$$

Dikdörtgensel bölgenin alanını bulmak için yukarıda gibi bir oran kullanan öğrenci seçeneklerde verilmiş olan bölgelerden hangisinin alanını bulurken bu yöntemle doğru sonuca ulaşır?

A) Paralelkenarsal bölge
B) İkizkenar yamuksal bölge
C) Düzgün altıgensel bölge
D) Dik yamuksal bölge
E) Beşgensel bölge

Gerekçe:...
....

Soru ile ilgili öneri:..
....
....

24)

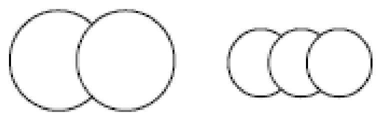
Boyutları 2,5m x 130cm x80 cm olan bir kutuya çevresi yaklaşık 75-78 cm olan basketbol toplarından kaç tane sığdırabilirsiniz?

A) 50 B) 80 C) 110 D) 160 D) 200

Gerekçe:...
....

Soru ile ilgili öneri:..
....
....

25)



Eşit uzunluktaki iki ipten ilkini iki eşit, ikincisini üç eşit parçaya bölerek Şekil I ve Şekil II'deki gibi çemberler oluşturulmuştur.

Sınıfta çember ve daire konusunu işleyen öğretmen öğrencilerinden Şekil I ve Şekil II'deki çemberlerin toplam çevreleri ile kapladıkları toplam alanları karşılaştırmalarını istiyor.

Aşağıda verilen öğrenci yanıtlarından hangisi doğrudur?

A) ***Öğrenci I***: Şekil I ve Şekil II'deki alanlar toplamı ve çevreler toplamı eşittir.

B) ***Öğrenci II***: Şekil I ve Şekil II'deki alanlar toplamı eşit, fakat Şekil I'deki çevreler toplamı daha büyüktür.

C) ***Öğrenci III***: Şekil I ve Şekil II'deki alanlar toplamı eşit, fakat Şekil I'deki çevreler toplamı daha küçüktür.

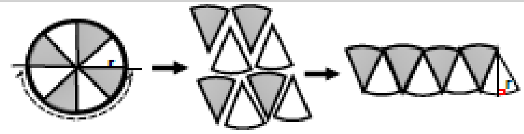
D) ***Öğrenci IV***: Şekil I ve Şekil II'deki çevreler toplamı eşit, fakat Şekil I'deki alanlar toplamı daha büyüktür.

E) ***Öğrenci V***:Şekil I ve Şekil II'deki çevreler toplamı eşit, fakat Şekil I'deki alanlar toplamı daha büyüktür.

Gerekçe:...
....

Soru ile ilgili öneri:..
....
....

26)



Şekil I *Şekil II* *Şekil III*

7.Sınıf öğrencilerine Öğrencilerine dairenin alanını anlatmayı planlayan bir öğretmen daha önceden hazırlamış olduğu Şekil I'deki daireyi Şekil II'deki gibi eş dilimler halinde keser ve Şekil III'teki gibi yeniden düzenler.

Etkinlikte paralelkenarsal bölgenin alan formülünü kullanarak dairenin alan formülünü elde ettirmeye çalışmaktadır. Fakat sınıftaki bir öğrenci oluşan yeni şeklin tabanları eğrilerden oluştuğu için paralelkenarsal bölgenin alan formülünün kullanılamayacağını iddia ediyor.

Bu ilişkilendirmeyi anlamakta güçlük çeken bir öğrencinin aşağıdaki durumların hangi ya da hangilerinde benzer şekilde zorluk yaşaması beklenir?

I) Bir silindirin içine yerleştirilenaynı yarıçaplı hamurdan yapılmış kürenin ezilerek, silindirin hacim formülünden kürenin hacim formülünün elde edilmesi

II) Eş büyüklüklerde oluşturulmuş konilerden küre oluşturularak koninin hacim formülünden kürenin hacim formülünün elde edilmesi

III) Hazırlanan eş büyüklüklerdeki daire dilimleri ile kürenin kaplanarak dairenin alan formülünden kürenin yüzey alanı formülünün elde edilmesi

IV) Kürenin içini yüksekliği eş bir koni yardımı ile sıvıyla doldurup daha sonra koninin hacim formülünden kürenin hacim formülünün elde edilmesi

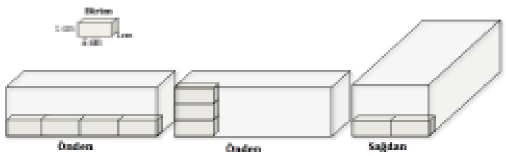
A) I, II, III, IV B) I, IV C) II, III
D) Yalnız II E) Yalnız III

Gerekçe:...
....

Soru ile ilgili öneri:..
....
....

İletişim için:yesen@metu.edu.tr

27)



Sınıfta prizmalarda hacim konusunu işlemekte olan öğretmen, içine birimler yerleştirilmiş şekilde önden ve sağdan görünümü verilen dikdörtgenler prizmasının hacmini soruyor.

Bir öğrenci prizmanın hacmini 4x2x3= 24 br^3 olarak buluyor.

Öğrencinin yanıtı ile ilgili olarak aşağıda verilen ifadelerden hangisi doğrudur?

A) Prizmanın içine yerleştirilen birimin metrik ölçüleri verildiği için hacim cm^3cinsinden bulunmalıdır.

B) Öğrenci yanıtı br^3 olarak bulduğu için sonuç doğrudur. Prizmaların hacimlerini küp dışındaki birimlerle bulurken birim sayılarını kullanarak da bulabiliriz.

C) Öğrenci tesadüfen doğru sonuç bulmuştur.

D) Birim sayılarını çarparak doğru sonuç elde etmek mümkün değildir.

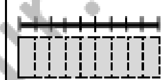
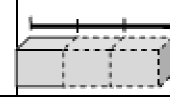
E) Öğrencinin yanıtı yanlıştır.

Gerekçe:...
....

Soru ile ilgili öneri:..
....
....

28)

Sınıfta uzunlukları ölçme konusunu işleyen bir öğretmen öğrencilerine sabit boyda çubuklar ve farklı büyüklükte birimler dağıtarak öğrencilerinden çubukların uzunluklarını dağıttığı birimlerle ölçmelerini istiyor.

	Durum I	Durum II
Öğrenci I: Her ikisi için de aynı araç kullanıldığı için uzunluk ölçüleri aynıdır.		
Öğrenci II: Her iki uzunluk için 3 birim kullanıldığı için uzunluk ölçüleri aynıdır.		
Öğrenci III: Her iki uzunluk için farklı sayıda birim kullanıldığı için uzunluk ölçüleri farklıdır.		
Öğrenci I IV: Her iki ölçümde farklı birimler kullanıldığı için uzunluk ölçüleri farklıdır.		

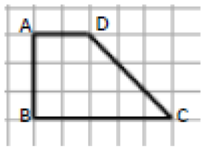
Yukarıda yanıtları verilen öğrencilerden hangi ikisi uzunluk ölçme konusunda benzer düşünceye sahiptir?

A) I-II B) I-III C) II-III D) II-IV E) III-IV

Gerekçe:...
....

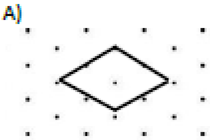
Soru ile ilgili öneri:..
....
....

29)

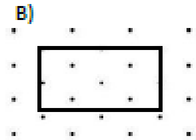


Yanda verilen şeklin çevresini 13 br olarak hesaplayan bir öğrenci için aşağıdaki örneklerden hangisiöğrencinin yaptığı hatayı fark ettirmeye yöneliktir?

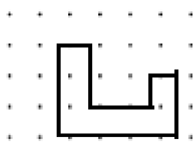
A)



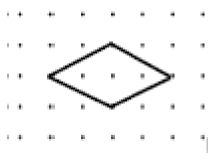
B)



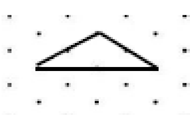
C).



D)



E)



Gerekçe:...
....

Soru ile ilgili öneri:...
....
....

TESTİNİZ BİTMİŞTİR.

ÇALIŞMAYA KATILDIĞINIZ İÇİN ÇOK TEŞEKKÜR EDERİM.

APPENDIX B

FINAL FORMS OF ITEMS

İletişim için:yesen@metu.edu.tr

Test_4

ODTÜ

PEDAGOJİK ALAN BİLGİSİ TESTİ (ÖLÇME)

ADI :..

SOYADI :...

CİNSİYET : KADIN ☐ ERKEK: ☐

ÜNİVERSİTE :..............................BÖLÜM : ....

GENEL NOT ORTALAMASI : ....

DÖNEM : ....

GENEL AÇIKLAMA

1.Bu sınavda her katılımcıya bir soru kitapçığı verilecektir.

2.Bu soru kitapçığı her biri **5 seçenekten** oluşan **15 sorudan** oluşmaktadır.

4.Testin tümü için öngörülen cevaplama süresi **25 Dakika**'dır.

5.Cevaplamaya istediğiniz sorudan başlayabilirsiniz. Cevaplarınızı kitapçık üzerine işaretlemeyi unutmayınız.

5.Bu ölçeğin sonuçları sadece araştırma amaçlı kullanılacak olup hiçbir şekilde notlarınızı **etkilemeyecektir.**

Teşekkürler.

İLETİŞİM BİLGİLERİ
Yasemin ESEN
ODTÜ Eğitim Fakültesi
İlköğretim Matematik Eğitimi Bölümü
06531 Çankaya- ANKARA
e-mail: yaseminesen@gmail.com

Tel: 0 312 210 40 65

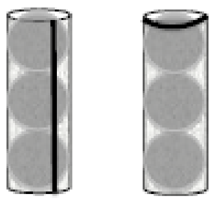
- 1 -

İletişim için:yesen@metu.edu.tr

SORULAR

1)

Aşağıdaki şekillerde içlerine üçer tane özdeş topsığdırılabilen iki özdeş silindir verilmiştir.



Şekil I *Şekil II*

Bu kutulardaki şeritlerden birisi *Şekil I*'deki gibi dikine, diğeri ise *Şekil II*'deki gibi kapağın çevresine yerleştirilmiştir.

Bir öğretmen, öğrencilerinden bu iki şeridin uzunluğunu karşılaştırmalarını istiyor.

Bu duruma göre aşağıda verilen öğrenci yanıtlarından hangisi doğrudur?

A) Şekil I'deki şerit, Şekil II'deki şeritten daha kısadır.

B) Şekil I'deki şerit, Şekil II'deki şeritten daha uzundur.

C) Şekil I ve Şekil II'deki şeritlerin boyları eşittir.

D) Sayısal bir değer olmadığı için bir şey söylenemez.

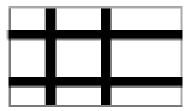
E) Yeterince büyük toplar kullanıldığında Şekil II'deki şerit, Şekil I'deki şeritten daha uzun olur.

Yukarıdaki soruya verdiğim yanıtın doğruluğundan:

Hiç emin değilim ○ Emin değilim ○ Kararsızım ○ Eminim ○ Çok eminim ○

2)

Alan konusunu işleyen bir öğretmen öğrencilerine aşağıdaki soruyu soruyor.



Kenar uzunlukları 200 m ve 100 m olan dikdörtgen şeklindeki bir arsa, kenarlarına dik olan 2 m genişliğinde yollar açılarak şekildeki gibi 9 parsele ayrılmıştır. Açılan yolların toplam alanı kaç m^2 dir? *(DPY/ 2005 – 7. Sınıf Soru 21)*

Soruya 1200 m^2 yanıtını veren öğrencinin, aynı arsanın, aynı genişlikteki 3 yatay ve 3 dikey yol ile 16 parsele ayrılması durumunda yolların toplam alanını kaç m^2 bulması beklenir?

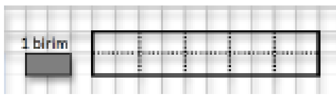
A) 1796 B) 1791 C) 1788 D) 1764 E) 1800

Yukarıdaki soruya verdiğim yanıtın doğruluğundan:

Hiç emin değilim ○ Emin değilim ○ Kararsızım ○ Eminim ○ Çok eminim ○

İletişim için:yesen@metu.edu.tr

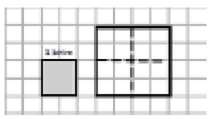
3)

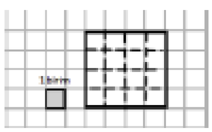


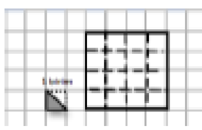
Bir öğretmen, öğrencilerinden yukarıdaki dikdörtgensel bölgenin alanını yanında verilen birim cinsinden ölçmelerini istiyor.

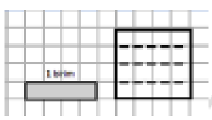
Bir öğrenci, soruyu hatalı olarak20 birim olarak yanıtlıyor.

Öğrencinin aşağıda verilen örneklerden hangiya da hangilerinde benzer şekilde hata yapması beklenir?

I) 

II) 

III) 

IV) 

A) Hepsi B) I, II, IV C) I, III, IV

D) I, IV E) Yalnız III

Yukarıdaki soruya verdiğim yanıtın doğruluğundan:

Hiç emin değilim ○ Emin değilim ○ Kararsızım ○ Eminim ○ Çok eminim ○

4)

MEB 6. Sınıf Matematik Öğretim Programındaki uyarılar dikkate alındığında, aşağıda verilen sorulardan hangisi ya da hangileri 6. Sınıf seviyesinde verilen *"Dikdörtgenler prizmasının yüzey alanını hesaplar."* **kazanımını <u>doğrudan ölçmeye</u> yöneliktir?**

I. Ayrıt uzunlukları 30 cm,40 cm ve 20 cm olan kutunun bütün yüzleri ayna ile kaplanmak istenirse, kaç cm^2 aynaya ihtiyaç olur?

II. Yüzey alanı 150 cm^2 olan bir dikdörtgenler prizmasının, iki ayrıtının uzunluğu 10 cm ve 3 cm ise diğer ayrıtının uzunluğu kaç cm'dir?

III. Ayrıt uzunlukları a,b, c olan dikdörtgenler prizmasının yüzey alan formülünün " ***2(a.b) +2(b.c) + 2 (a.c)*** " olduğunu gösterin.

A) Yalnız I
B) Yalnız II
C) Yalnız III
D) II, III
E) I, II, III

Yukarıdaki soruya verdiğim yanıtın doğruluğundan:

Hiç emin değilim ○ Emin değilim ○ Kararsızım ○ Eminim ○ Çok eminim ○

İletişim için:yesen@metu.edu.tr

5)

Sınıfta alan konusunu işleyen bir öğretmen, öğrencilerinden aşağıda <u>farklı büyüklükteki birimler</u> cinsinden değerleri verilen geometrik şekillerin alanlarını karşılaştırmalarını istiyor.

	Şekil I	*Şekil II*
Değer:	25 birim kare	30 birim üçgen

Öğrenci I: Şekil II, Şekil I'den büyüktür.

	Şekil I	*Şekil II*
Değer:	26 birim kare	15 birim kare

Öğrenci II: Şekil I, Şekil II'den büyüktür.

	Şekil I	*Şekil II*
Değer:	12 cm^2	9 cm^2

Öğrenci III: Şekil I, Şekil II'den büyüktür.

Yukarıdakisayısal bilgilere dayanarak ifade edilen öğrenci önermelerinden hangisi ya da hangileri <u>her zaman</u> doğru değildir?

A) I, II, III
B) I, II
C) Yalnız I
D) Yalnız II
E) Yalnız III

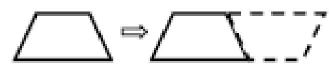
Yukarıdaki soruya verdiğim yanıtın doğruluğundan:

Hiç emin değilim	Emin değilim	Kararsızım	Eminim	Çok eminim
◯	◯	◯	◯	◯

6)

Yamuksal bölgenin alanını işlemekte olan bir öğretmen, öğrencilerine ikizkenar bir yamuk modeli verip, öğrencilerinden yamuksal bölgenin alanını bulmaya yönelik farklı stratejiler geliştirmelerini istemiştir.

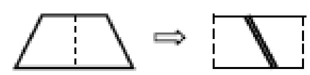
Öğrencilerden gelen aşağıdakistratejilerden hangi ya da hangileri <u>herhangi bir</u>yamuksal bölgenin alanını bulmaya yönelik olarak genellenebilir?

I) 

Şeklin eşini şekildeki gibi ekleyerek paralelkenarsal bölgenin alan formülünü kullanıp, yarısını almak.

II) 

Şekilde verildiği gibi yamuğu iki üçgensel bölgeye ayırıp, üçgensel bölgenin alan formülünü kullanıp, alanlarını toplamak.

III) 

Kesik çizgili yerden kesip, şekildeki gibi birleştirerek dikdörtgensel bölgenin alan formülünü kullanmak.

A) I, II, III B) II, III C)I, II

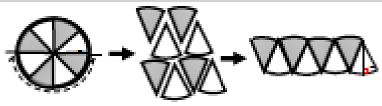
D) Yalnız II E) Yalnız III

Yukarıdaki soruya verdiğim yanıtın doğruluğundan:

Hiç emin değilim	Emin değilim	Kararsızım	Eminim	Çok eminim
◯	◯	◯	◯	◯

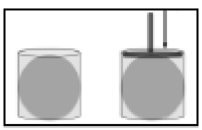
İletişim için:yesen@metu.edu.tr

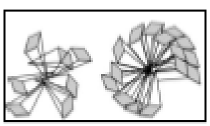
7)

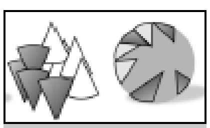


Öğrencilerine dairenin alanını anlatmayı planlayan bir öğretmen daha önceden hazırlamış olduğu daireyi eş dilimler halinde keser ve yukarıdaki gibi yeniden düzenler.
Paralelkenarsal bölgenin alan formülünü kullanarak dairenin alan formülünü elde ettirmeye çalışır.
Sınıftaki bir öğrenci oluşan yeni şekil eğrilerden oluştuğu için paralelkenarsal bölgenin alan formülünün kullanılamayacağını iddia eder.

Etkinlikteki matematiksel yaklaşımı anlamakta güçlük çeken bu öğrencinin aşağıdaki durumların hangi ya da hangilerinde benzer şekilde düşünüp öğrenme güçlüğü yaşaması beklenir?

I) 
Bir silindirin içinde yarıçapı eşit oyun hamurundan yapılmış bir kürenin ezilerek, silindirin hacim formülünden kürenin hacim formülünün elde edilmesi.

II) 
Eş büyüklüklerde oluşturulmuş kare piramitlerden küre oluşturularak piramitin hacim formülünden kürenin hacim formülünün elde edilmesi.

III) 
Kürenin yüzeyinin eş büyüklükteki daire dilimleri ile kaplanarak, dairenin alan formülünden kürenin yüzey alanı formülünün elde edilmesi

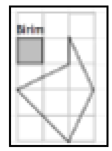
A) I, II, III B) II, III C)I, III D) Yalnız II E) Yalnız I

Yukarıdaki soruya verdiğim yanıtın doğruluğundan:

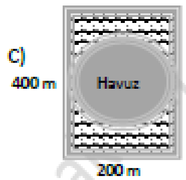
Hiç emin değilim	Emin değilim	Kararsızım	Eminim	Çok eminim
○	○	○	○	○

8) **MEB Matematik Öğretim Programı dikkate alındığında, alan konusu ile ilgili aşağıdaki verilen sorulardan hangisi 6. Sınıf seviyesine <u>uygun değildir</u>?**

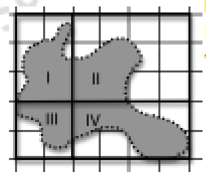
A) 
Yandaki haritada taralı bölge 200 hektardır. Bu alan kaç dönümdür?

B) 

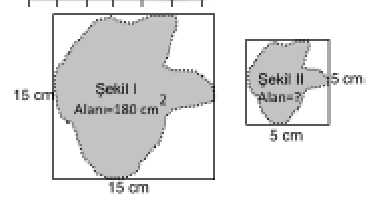
Yandaki şeklin alanı kaç birimdir?

C) 

Yandaki krokide çapı 200 m olan havuz dışında kalan kaç m^2' dir?

D) 

Yanda verilen taralı bölgenin tamamı $36m^2$' dir. IV numaralı bölgenin alanı yaklaşık kaç m^2' dir?

E) 

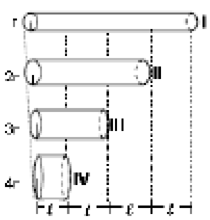
Şekilde verilen değerlere göre Şekil II'nin alanı kaç cm^2'dir?

Yukarıdaki soruya verdiğim yanıtın doğruluğundan:

Hiç emin değilim	Emin değilim	Kararsızım	Eminim	Çok eminim
○	○	○	○	○

İletişim için:yesen@metu.edu.tr

9)



İçi boş r,2r,3r ve 4r yarıçaplı silindirlerin boyları şekildeki gibidir. Silindirlerin her birisi yarıya kadar su ile doldurulduğu zaman aldıkları su miktarlarını sıralayınız.

(DPY 2005/9-10-11. Sınıf Soru

Soruya ***"I < II < III < IV"*** gibi yanıt veren bir öğrenci silindirin hacmine etki eden değişkenlerden sadece birini dikkate alarak sıralama yapmıştır.

Öğrencinin yaşadığı bu güçlüğü aşması için aşağıda verilen etkinliklerden hangisi kullanılabilir?

A) Soruda yatay şekilde verilen silindirleri dikey hale getirdikten sonra hacimlerini karşılaştırmak.

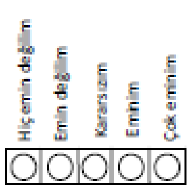
B) Soruda yarısı doldurulan silindirlerin tamamını doldurduktan sonra hacimlerini karşılaştırmak.

C) Yükseklikleri aynı taban alanları farklı silindirlerin hacimlerini karşılaştırmak.

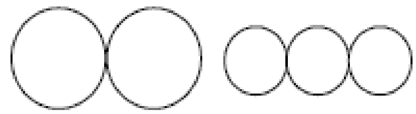
D) Yükseklikleri farklı taban alanları aynı silindirlerin hacimlerini karşılaştırmak.

E) Soruda verilen silindirlerin açınımlarını çizerek hacimlerini karşılaştırmak.

Yukarıdaki soruya verdiğim yanıtın doğruluğundan:



10)



Şekil I Şekil II

Eşit uzunluktaki iki ipten ilkini iki eşit, ikincisini üç eşit parçaya bölerek Şekil I ve Şekil II'deki gibi çemberler oluşturulmuştur.

Sınıfta çember ve daire konusunu işleyen öğretmen öğrencilerinden Şekil I ve Şekil II'deki çemberlerin toplam çevreleri ile çevreledikleri toplam alanları karşılaştırmalarını istiyor.

Aşağıda verilen öğrenci yanıtlarından hangisi doğrudur?

A) Şekil I ve Şekil II'deki alanlar toplamı ve çevreler toplamı eşittir.

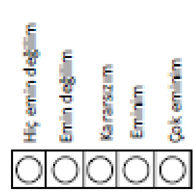
B) Şekil I ve Şekil II'deki alanlar toplamı eşit, fakat Şekil I'deki çevreler toplamı daha büyüktür.

C) Şekil I ve Şekil II'deki alanlar toplamı eşit, fakat Şekil I'deki çevreler toplamı daha küçüktür.

D) Şekil I ve Şekil II'deki çevreler toplamı eşit, fakat Şekil I'deki alanlartoplamı daha büyüktür.

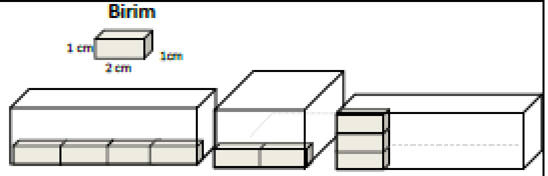
E) Şekil I ve Şekil II'deki çevreler toplamı eşit, fakat Şekil I'deki alanlar toplamı daha küçüktür.

Yukarıdaki soruya verdiğim yanıtın doğruluğundan:



İletişim için:yesen@metu.edu.tr

11)



Prizmalarda hacim konusunu işlemekte olan öğretmen, içine birimler yerleştirilmiş şekilde önden ve sağdan görünümü verilen dikdörtgenler prizmasının hacmini soruyor.

Bir öğrenci prizmanın hacmini *4x2x3=24 Birim*olarak buluyor.

Öğrencinin bulduğu sonucun hatalı olduğu biliniyorsa, bu hatanın kaynağı aşağıdakilerden hangisi olabilir?

A) Kullanılan birimin metrik ölçüleri verilmiştir. Sonuç cm^3 cinsinden bulunmalıdır.

B) Kullanılan birim, kare prizmadır. Hacim hesaplanırken kullanılan birim "küp" olmalıdır.

C) Hacim ifade edilirken " 24 Birim" yerine "24 $Birim^3$" kullanılmalıdır.

D) Birimler yanlış yerleştirilmiştir. Bu bağlı hesaplama sonucu yanlıştır.

E) Cisim dikdörtgenler prizması iken, birim kare prizmadır. İçine yerleştirilen birimle cismin şekli aynı olmalıdır.

Yukarıdaki soruya verdiğim yanıtın doğruluğundan:

Hiç emin değilim	Emin değilim	Kararsızım	Eminim	Çok eminim
○	○	○	○	○

12)

Yanda verilen tuvalet kâğıdının sarıldığı karton silindirin yarıçapı 2 cm ve sarılan tuvalet kâğıdının kalınlığı 6 cm' dir.

Kâğıdın ruloya 200 kez dolandığı düşünülürse, tuvalet kâğıdının uzunluğu <u>yaklaşık</u> kaç metredir?

A) 24 B)36 C)60 D) 72 E)96

Yukarıdaki soruya verdiğim yanıtın doğruluğundan:

Hiç emin değilim	Emin değilim	Kararsızım	Eminim	Çok eminim
○	○	○	○	○

13)

Koninin yüzey alan bağıntısını öğrencilerle birlikte sınıfta elde edecek bir öğretmen, konuya başlamadan önce öğrencilerinde aşağıda verilen kavramlardan hangi ya da hangilerinde bilgi eksikliği olmadığından emin olmalıdır?

I. Dairenin alanı
II. Daire diliminin alanı
III. Çemberin çevresi
IV. Yükseklik

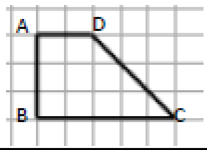
A) I, II, III, IV B) I, II, III C) I, II

D) I, IV E) Yalnız II

Yukarıdaki soruya verdiğim yanıtın doğruluğundan:

Hiç emin değilim	Emin değilim	Kararsızım	Eminim	Çok eminim
○	○	○	○	○

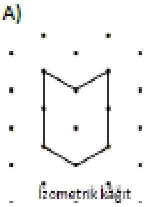
14)



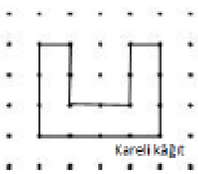
Bir öğrenci yanda verilen dörtgenin çevresini 13 br olarak hesaplamıştır.

Aşağıdaki örneklerden hangisinde öğrenci benzer bir hata yapar?

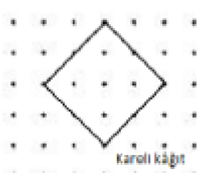
A)



B)

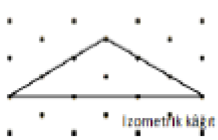


C)



D)

E)



Yukarıdaki soruya verdiğim yanıtın doğruluğundan:

Hiç emin değilim	Emin değilim	Kararsızım	Eminim	Çok eminim
◯	◯	◯	◯	◯

15)

Ölçme kavramını işleyen bir öğretmen dersine resimdeki gibi bir kova göstererek başlıyor ve sınıfta öğrencileri ile birlikte aşağıdaki etkinlikleri yapıyor.

Öğretmenin sınıfta yapmış olduğu etkinliklerden hangisi <u>kovanın bir özelliğine</u> ilişkin ölçüm <u>değildir</u>?

A) Kovanın dış yüzeyini alanı yaklaşık 25 cm^2 olan kare şeklindeki kağıtlarla kaplayıp kaç parça yapıştırıldığını saymak.

B) Kovanın ağzındaçevresi yaklaşık 9 cm olan madeni para yuvarlayıp paranın kaç tur yaptığını saymak.

C) Su içinde saniyede 2 metre sabit hızla çökebilen bir cihazı, su dolu kova içine attıktan sonra tabana ulaşana kadar geçen süreyi ölçmek.

D) Kovanın içini doldurduktan sonra esneme katsayısı $(^{3\ cm}/_{10\ kg})$ olan esnek bir iple asıp ipteki esneme miktarını ölçmek.

E) Kovayı yarım litrelik pet şişelerdeki sularla doldurup kaç pet şişe boşaltıldığını saymak.

Yukarıdaki soruya verdiğim yanıtın doğruluğundan:

Hiç emin değilim	Emin değilim	Kararsızım	Eminim	Çok eminim
◯	◯	◯	◯	◯

APPENDIX C

INTERVIEW PROTOCOL FOR PRESERVICE TEACHERS

GÖRÜŞME SORULARI

1. Sınavla ilgili genel olarak dikkatini çeken noktalar oldu mu?
    a. Olduysa nelerdir?
    b. Sınav hakkında genel olarak ne düşünüyorsun?
2. Anlamakta güçlük çektiğin soru (lar) oldu mu?
    a. Ne açıdan güçlük çektin?
    b. Nasıl olsa daha anlaşılabilir olur?

3. Sınav sorularını kısaca bir hatırladıktan sonra,
    a. Senin için kolay olan soruları belirleyebilir misin?
        i. Neden kolay olduklarını düşünüyorsun?
    b. Senin için zor olan soruları belirleyebilir misin?
        i. Ne açıdan zorlandın?

4. İkilemde kaldığın sorular oldu mu?
    a. Oldu ise hangi (leri) ?
    b. İkilemde kalmanın sebebi nedir?

5. Testte gereksiz bulduğun soru var mı?
    a. Ne açıdan gereksiz olduğunu düşünüyorsun?

6. Seçenekler yardımı ile sonuca ulaştığın sorular var mı?
    a. Varsa hangi (leri) ?
    b. Seçenekler sana nasıl yardım etti?

7. Sonucunun olmadığını düşündüğün soru (lar) var mı?
    a. Hangileri? Neden sonucunun olmadığını düşünüyorsun?

8. Genel olarak söylemek istediğin başka bir şey var mı?

APPENDIX D

INTERVIEW PROTOCOL FOR INSTRUCTORS

İsim:

Deneyim:

Ağırlıklı olarak hangi dersler:

Dersin programı paylaşım: Evet …… Hayır ……

ÖÖY I:

Konular ve işleniş:

ÖÖY II:

Konular ve işleniş:

Ders kitabı:

Neden:

Ölçme konusu işleniş:

Sınav soru örnekleri: Evet   ……   Hayır ……

Grupla ilgili paylaşmak istedikleriniz:

APPENDIX E

OBJECTIVES ON MEASUREMENT CONCEPTS FOR GRADES 6-8 IN TURKISH ELEMENTARY MATHEMATICS CURRICULUM

6. SINIF	**Açıları Ölçme**	Tümler, bütünler ve ters açıların ölçülerini hesaplar.
	Uzunlukları Ölçme	Uzunluk ölçme birimlerini açıklar ve birbirine dönüştürür. Düzlemsel şekillerin çevre uzunluklarını strateji kullanarak tahmin eder. Düzlemsel şekillerin çevre uzunlukları ile ilgili problemleri çözer ve kurar. Çokgenlerin kenar uzunlukları ile çevre uzunluğu arasındaki ilişkiyi açıklar.
	Alanı Ölçme	Alan ölçme birimlerini açıklar ve birbirine dönüştürür. Düzlemsel bölgelerin alanlarını strateji kullanarak tahmin eder. Düzlemsel bölgelerin alanları ile ilgili problemleri çözer ve kurar. Dikdörtgenler prizması, kare prizma ve küpün yüzey alanlarını hesaplar. Dikdörtgenler prizması, kare prizma ve küpün yüzey alanı ile ilgili problemleri çözer ve kurar.
	Zamanı Ölçme	Zaman ölçme birimleriyle ilgili problemleri çözer ve kurar.
	Hacmi Ölçme	Dikdörtgenler prizması, kare prizma ve küpün hacmine ait bağıntıları oluşturur. Dikdörtgenler prizması, kare prizma ve küpün hacmini strateji kullanarak tahmin eder. Dikdörtgenler prizması, kare prizma ve küpün hacmi ile ilgili problemleri çözer ve kurar. Hacim ölçme birimlerini açıklar ve birbirine dönüştürür.
	Sıvıları Ölçme	Sıvı ölçme birimlerini açıklar ve birbirine dönüştürür. Hacimölçmebirimleriilesıvıölçmebirimleriarasındailişkiyiaçıklar. Sıvı ölçme birimleri ile ilgili problemleri çözer ve kurar.
7. SINIF	**Doğrular ve Açılar**	Aynı düzlemde olan üç doğrunun birbirine göre durumlarını belirler (ve inşa eder). Yöndeş, iç, içters, dış ve dış ters açıları belirleyerek isimlendirir. açıklanması) Paralelikidoğrununbirkesenleyaptığıaçılarıneşolanlarınıvebütünlerolanlarınıbelirler.
	Çokgenler	Çokgenlerin köşegenlerini, iç ve dış açılarını belirler. Dörtgenlerin kenar, açı ve köşegen özelliklerini belirler.
	Eşlik ve Benzerlik	Çokgenlerikarşılaştırarakeşolupolmadıklarınıbelirlervebirçokgeneeşçokgenleroluşturur. Çokgenlerikarşılaştırarakbenzerolupolmadıklarınıbelirlervebirçokgenebenzerçokgenleroluşturur.

	Çember ve Daire	Çemberin özelliklerini belirler ve çember modeli inşa eder. Çemberin düzlemde ayırdığı bölgeleri belirler. Çember ile doğrunun ilişkisini belirler. Çember veya dairede merkez açı ve çevre açı ile bu açıların gördüğü yayları belirler. Aynıyayıgörenmerkezaçınınölçüsüileçevreaçınınölçüsüarasındakiiliş kiyibelirler.
	Geometrik Cisimler	Dairesel silindirin temel elemanlarını belirler, inşa eder ve açınımını çizer. Yüzlerinin farklı yönlerden görünümlerine ait çizimleri verilen yapıları, birim küplerle oluşturur ve izometrik kâğıda çizer.
	Dönüşüm Geometrisi	Yansımayı açıklar. Dönme hareketini açıklar. Düzlemdebirnoktaetrafındavebelirtilenbiraçıyagöreşekilleridöndürere kçiziminiyapar.
	Örüntü ve Süslemeler	Çokgensel bölge modelleriyle bir bölgeyi döşeyerek süsleme yapar. Düzgünçokgenselbölgemodelleriyleoluşturulansüslemelerdekikodları belirler. Yansıma, öteleme ve dönme hareketleri ile süsleme yapar.
8. SINIF	**Üçgenlerde Ölçme**	Üçgenlerde benzerlik şartlarını problemlerde uygular. Pythagoras (Pisagor) bağıntısını problemlerde uygular. Dik üçgendeki dar açıların trigonometrik oranlarını problemlerde uygular.
	Geometrik Cisimlerin Hacimleri	Dik prizmaların hacim bağıntılarını oluşturur. Dik piramidin hacim bağıntısını oluşturur. Dik dairesel koninin hacim bağıntısını oluşturur. Kürenin hacim bağıntısını oluşturur. Geometrik cisimlerin hacimleri ile ilgili problemleri çözer ve kurar. Geometrik cisimlerin hacimlerini strateji kullanarak tahmin eder.
	Geometrik Cisimlerin Yüzey Alanları	Dik prizmaların yüzey alanının bağıntılarını oluşturur. Dik piramidin yüzey alanının bağıntısını oluşturur. Dik dairesel koninin yüzey alanının bağıntısını oluşturur. Kürenin yüzey alanının bağıntısını oluşturur. Geometrik cisimlerin yüzey alanları ile ilgili problemleri çözer ve kurar. Geometrik cisimlerin yüzey alanlarını strateji kullanarak tahmin eder.

Summary of Commonly Addressed Measurement Concepts for Grades 6-8

6. SINIF	**Uzunlukları Ölçme**	Uzunluk ölçme birimlerini açıklar ve birbirine dönüştürür. Düzlemsel şekillerin çevre uzunluklarını strateji kullanarak tahmin eder. Düzlemsel şekillerin çevre uzunlukları ile ilgili problemleri çözer ve kurar. Çokgenlerin kenar uzunlukları ile çevre uzunluğu arasındaki ilişkiyi açıklar.
	Alanı Ölçme	Alan ölçme birimlerini açıklar ve birbirine dönüştürür. Düzlemsel bölgelerin alanlarını strateji kullanarak tahmin eder. Düzlemsel bölgelerin alanları ile ilgili problemleri çözer ve kurar. Dikdörtgenler prizması, kare prizma ve küpün yüzey alanlarını hesaplar. Dikdörtgenler prizması, kare prizma ve küpün yüzey alanı ile ilgili problemleri çözer ve kurar.
	Hacmi Ölçme	Dikdörtgenler prizması, kare prizma ve küpün hacmine ait bağıntıları oluşturur. Dikdörtgenler prizması, kare prizma ve küpün hacmini strateji kullanarak tahmin eder. Dikdörtgenler prizması, kare prizma ve küpün hacmi ile ilgili problemleri çözer ve kurar. Hacim ölçme birimlerini açıklar ve birbirine dönüştürür.
7. SINIF	**Çember ve Daire**	Çemberin özelliklerini belirler ve çember modeli inşa eder. Çemberin düzlemde ayırdığı bölgeleri belirler. Çember ile doğrunun ilişkisini belirler. Çember veya dairede merkez açı ve çevre açı ile bu açıların gördüğü yayları belirler. Aynıyayıgörenmerkezaçınınölçüsüileçevreaçınınölçüsüarasındakiilişkiyibelirler.
	Geometrik Cisimler	Dairesel silindirin temel elemanlarını belirler, inşa eder ve açınımını çizer. Yüzlerinin farklı yönlerden görünümlerine ait çizimleri verilen yapıları, birim küplerle oluşturur ve izometrik kâğıda çizer.
8. SINIF	**Üçgenlerde Ölçme**	Üçgenlerde benzerlik şartlarını problemlerde uygular. Pythagoras (Pisagor) bağıntısını problemlerde uygular. Dik üçgendeki dar açıların trigonometrik oranlarını problemlerde uygular.
	Geometrik Cisimlerin Hacimleri	Dik prizmaların hacim bağıntılarını oluşturur. Dik piramidin hacim bağıntısını oluşturur. Dik dairesel koninin hacim bağıntısını oluşturur. Kürenin hacim bağıntısını oluşturur. Geometrik cisimlerin hacimleri ile ilgili problemleri çözer ve kurar. Geometrik cisimlerin hacimlerini strateji kullanarak tahmin eder.
	Geometrik Cisimlerin Yüzey Alanları	Dik prizmaların yüzey alanının bağıntılarını oluşturur. Dik piramidin yüzey alanının bağıntısını oluşturur. Dik dairesel koninin yüzey alanının bağıntısını oluşturur. Kürenin yüzey alanının bağıntısını oluşturur. Geometrik cisimlerin yüzey alanları ile ilgili problemleri çözer ve kurar. Geometrik cisimlerin yüzey alanlarını strateji kullanarak tahmin eder.

APPENDIX F

THE COURSE LOAD OF PRE-SERVICE MATHEMATICS EDUCATION UNDERGRADRADUATE PROGRAM

İLKÖĞRETİM MATEMATİK ÖĞRETMENLİĞİ LİSANS PROGRAMI

I. YARIYIL

	DERSİN ADI	T	U	K
A	Genel Matematik	4	2	5
GK	Türkçe I: Yazılı Anlatım	2	0	2
GK	Atatürk İlkeleri ve İnkılap Tarihi I	2	0	2
GK	Bilgisayar I	2	2	3
GK	Yabancı Dil I	3	0	3
MB	Eğitim Bilimine Giriş	3	0	3
TOPLAM		**16**	**4**	**18**

II. YARIYIL

	DERSİN ADI	T	U	K
A	Soyut Matematik	3	0	3
A	Geometri	3	0	3
GK	Türkçe II: Sözlü Anlatım	2	0	2
GK	Atatürk İlkeleri ve İnkılap Tarihi II	2	0	2
GK	Yabancı Dil II	3	0	3
GK	Bilgisayar II	2	2	3
MB	Eğitim Psikolojisi	3	0	3
TOPLAM		**18**	**2**	**19**

III. YARIYIL

	DERSİN ADI	T	U	K
A	Analiz I	4	2	5
A	Lineer Cebir I	3	0	3
A	Fizik I	4	0	4
A	Seçmeli I	2	0	2
GK	Bilimsel Araştırma Yöntemleri	2	0	2
MB	Öğretim İlke ve Yöntemleri	3	0	3
TOPLAM		**18**	**2**	**19**

IV. YARIYIL

	DERSİN ADI	T	U	K
A	Analiz II	4	2	5
A	Lineer Cebir II	3	0	3
A	Fizik II	4	0	4
GK	Seçmeli I	3	0	3
MB	Öğretim Teknolojileri ve Materyal Tasarımı	2	2	3
TOPLAM		**16**	**4**	**18**

V. YARIYIL

	DERSİN ADI	T	U	K
A	Analiz III	3	0	3
A	Analitik Geometri I	3	0	3
A	İstatistik ve Olasılık I	2	2	3
A	Cebire Giriş	3	0	3
GK	Bilim Tarihi*	2	0	2
MB	Seçmeli I	2	0	2
MB	Özel Öğretim Yöntemleri I	2	2	3
TOPLAM		**17**	**4**	**19**

VI. YARIYIL

	DERSİN ADI	T	U	K
A	Diferansiyel Denklemler	4	0	4
A	Analitik Geometri II*	3	0	3
A	İstatistik ve Olasılık II*	2	2	3
A	Özel Öğretim Yöntemleri II	2	2	3
GK	Türk Eğitim Tarihi*	2	0	2
GK	Topluma Hizmet Uygulamaları	1	2	2
MB	Ölçme ve Değerlendirme	3	0	3
TOPLAM		**17**	**6**	**20**

VII. YARIYIL

	DERSİN ADI	T	U	K
A	Elemanter Sayı Kuramı*	3	0	3
A	Seçmeli II	3	0	3
GK	Matematik Tarihi*	2	0	2
MB	Rehberlik	3	0	3
MB	Okul Deneyimi	1	4	3
MB	Sınıf Yönetimi	2	0	2
MB	Özel Eğitim*	2	0	2
TOPLAM		**16**	**4**	**18**

VIII. YARIYIL

	DERSİN ADI	T	U	K
A	Matematik Felsefesi*	2	0	2
GK	Seçmeli II	3	0	3
MB	Türk Eğitim Sistemi ve Okul Yönetimi	2	0	2
MB	Öğretmenlik Uygulaması	2	6	5
MB	Seçmeli II	3	0	3
TOPLAM		**12**	**6**	**15**

	Teorik	Uygulama	Kredi	Saat
GENEL TOPLAM	130	32	146	162

A: Alan ve alan eğitimi dersleri, **MB:** Öğretmenlik meslek bilgisi dersleri, **GK:** Genel kültür dersleri

APPENDIX G

TEZ FOTOKOPİSİ İZİN FORMU

ENSTİTÜ

Fen Bilimleri Enstitüsü ☐

Sosyal Bilimler Enstitüsü X

Uygulamalı Matematik Enstitüsü ☐

Enformatik Enstitüsü ☐

Deniz Bilimleri Enstitüsü ☐

YAZARIN

Soyadı : ESEN

Adı : YASEMİN

Bölümü : ELE

TEZİN ADI (İngilizce) : DEVELOPMENT OF A TEST FOR ASSESSING TEACHERS' MATHEMATICAL CONTENT KNOWLEDGE FOR TEACHING GEOMETRIC MEASUREMENT AT ELEMENTARY GRADE LEVEL

TEZİN TÜRÜ : Yüksek Lisans ☐ Doktora X

1. Tezimin tamamından kaynak gösterilmek şartıyla fotokopi alınabilir. ☐
2. Tezimin içindekiler sayfası, özet, indeks sayfalarından ve/veya bir bölümünden kaynak gösterilmek şartıyla fotokopi alınabilir. ☐
3. Tezimden bir bir (1) yıl süreyle fotokopi alınamaz. X

TEZİN KÜTÜPHANEYE TESLİM TARİHİ:

APPENDIX F

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: ESEN, Yasemin

Nationality: Turkish (TC)

Date and Place of Birth: 15 April 1980, Burdur

Marital Status: Single

Phone: +90 530 600 60 39

email: esenyasemin@gmail.com

EDUCATION

Degree	**Institution**	**Year of Graduation**
MS	METU - SSME	2005
BS	METU - Mathematics	2004
High School	Burdur Cumhuriyet High School	1998

WORK EXPERIENCE

Year	**Place**	**Enrollment**
2007- Present	Middle East Technical University, Elementary Mathematics Education	Research Assistant

FOREIGN LANGUAGES

English

APPENDIX H

TURKISH SUMMARY

TÜRKÇE ÖZET

İLKÖĞRETİM MATEMATİK ÖĞRETMENLERİNİN GEOMETRİK ÖLÇME KAVRAMLARINI ÖĞRETME BİLGİLERİNİ ÖLÇMEYE YÖNELİK TEST GELİŞTİRME

BÖLÜM 1

GİRİŞ

İlgili alan yazını 1850'li yıllardan bu yana öğretmenlerin bilgileriniölçmek için ölçme araçlarının kullanıldığını söylemektedir. Ölçme araçlarının içerik ve formatları kadar ölçme değerlendirme yöntemleri de bu süreçte teorik çerçevelerin doğrultusunda şekillenmiştir. Başlarda kısa sınırlı ve ağırlıklı olarak kişisel görüşmelerden oluşan ölçme değerlendirme formatları daha sonraki süreçte nitel, öğretmenlik özelliklerine yönelik olarak tasarlanmış daha kapsamlı yöntemlere yerini bırakmıştır.Eğitim öğretim sürecini etkileyen faktörler arasında belki de en önemli yere sahip olan faktör öğretmenlerdir (Fullan, 2000). Bu noktada öğretmenlerin hangi özelliklerinin eğitim öğretim sürecini daha fazla etkilediğini belirlemek önem kazanmaktadır.

Yaklaşık 1960'lı yıllardan beri üzerinde yoğunlukla çalışılanöğretmenlik bilgisi kavramı yıllar boyunca hem teori bakımından hem de araştırma metotları bakımından çok fazla değişiklik geçirmiştir. Her ne kadar öğretmenlerin sahip olmaları gereken yeterlikler teknik ayrıntılarıyla davranışlar olarak ifadelendirilmiş olsa da,Shulman (1986) çalışmasında o döneme kadar yapılan çalışmaların öğretmenlerin gözlenebilir özellikleri üzerine kurulu olmasına karşı çıkmış ve öğretmenlik üzerine yapılan çalışmaların öğretmenlerin düşünme yapılarından bağımsız olmasını eleştirmiştir. Hatta o güne kadar yapılan çalışmalardaki öğretmenin bilişsel yapısını göz ardı eden bu yaklaşımı "kayıp paradigma (*missingparadigm*) " olarak değerlendirmiştir. Aynı şekilde Leinhardt, 1990 yılındaki çalışmasında öğretmenlik bilgisini öğretmenlerin gözlemlenebilir belli davranışlar, öğrencilerin notları ya da öğretmenlerin sırf alana yönelik belli sınavlardan aldıkları notlara göre değerlendirilmesinin hatalı sonuçlar üretebileceğine vurgu yapmıştır.

Shulman bu gözlemlerin üzerine 1986 yılındaki çalışmasında *öğretmenlik bilgisini* sınıflandırdığı yeni bir model ve varsayımlar üzerine kurduğu öğretmenlik bilgisinin alt alanlarını sunmuştur.Bu çalışma ve sonraki çalışmalarında (Shulman, 1987) öne sürülen öğretmenlik bilgisi modelitamamen teorik ve varsayımlar üzerine kuruludur. Öğretmenlik bilgisinin başlangıçtaki ağırlıklı teorik yapısı, sonrasında yapılan deneysel çalışmalarla desteklenmiş ve desteklenmeye devam edilmektedir. Örneğin, matematik öğretmenin sahip olması gereken matematik bilgisinin, matematik alanında çalışan herhangi bir kişiden farklı olması gerektiğini söyleyen çalışmaların sayısı her geçen gün artmaktadır (Ma, 1999; Ball, 1993; Ball ve McDiarmid, 1990; Ball, 1991; Ball, Lubienski ve Mewborn, 2001; Ball ve Bass, 2003; Mason ve Spence, 1999; Stylianides ve Ball, 2008).

Bu noktayı daha açıkça belirtmek gerekirse, çok temel matematik bilgisine sahip olan herhangi bir yetişkin $\sqrt{3^2+4^2}$ =$\sqrt{5^2}$=5 işlemini kolaylıkla sonuçlandırabilir.Fakat sınıf ortamlarında sıkça rastlanan$\sqrt{3^2+4^2}$=$\sqrt{3^2}$+$\sqrt{4^2}$ = 3+4 =7şeklindeki öğrenci hatasının nereden kaynaklandığını bulma, öğrencinin yaptığı bu yanlışla ilgili karşıt örnekler oluşturabilme, öğrencinin oluşturduğu yanlış çözümstratejisi yerine doğrusunu oluşabilmesi için uygun yöntemin belirlenmesigibi gerekli öğretmenlik bilgilerinin her biri pedagojik alan bilgisinin ayrı alt bilgi alanlarını oluşturmaktadır (Ball, ve diğer, 2001). Öğrencinin ulaşmış olduğu yanlış sonucu farketmenin yanısıra sorunun nerden kaynaklandığını bulmak, öğrencinin yanlış yapmasına neden olan düşünce sistemini farketmek, öğrencinin cevabının yanlış olduğu konusunda onu ikna etmek, matematik öğretmenin sahip olması gereken becerilerdendir. Hatta bu noktada öğretmenlerin konu ile ilgili öğrencilerin düştükleri genel kavram yanılgılarını bilmeleri, karşılaşabilecekleri işlem hataları ve zorluklardan haberdar olmaları gerekmektedir. Kısaca öğretmenlik bilgisi öğrencilerdeki temel matematiksel becerileri edindirmenin yanı sıra karşılaşılan bir durumla ilgili öğrencilerin yaşadıkları zorluk ve problemlerin kaynaklarını tespit etme, problemlerle ilgili çözüm yöntemlerini bilme gibi derin ve bütüncül bilgi gerektirmektedir.Öğretmenlik bilgisi ve alt alanlarınıntanımlanmasının ardından öğretmenlerin ve öğretmen adaylarının değerlendirmeleri amacıyla öğretmenlik bilgisine yönelik ölçme araçlarının kullanılması ve bu amaca yönelik çalışmaların yapılması bir gereklilik haline gelmiştir.

1.1. Çalışmanın Amacı

Bu çalışma kapsamında ilköğretim matematik öğretmen adaylarının uzunluk, çevre, alan ve hacim ölçme konularında öğretim yeterliliklerini yönelik bir test geliştirilmesi ve bu testin geçerlik güvenirlik çalışmaları yapılması hedeflenmiştir.

1.2. Araştırma Sorusu

Geliştirilen Ölçme Kavramlarını Öğretme Bilgi Testi (ÖKÖBT) ne ölçüde geçerlidir?

1.3. Çalışmanın Önemi

Öğretmenlik Bilgisi üzerine çalışma yapan araştırmalar pedagojik alan bilgisinin sadece var olmadığını aynı zamanda öğretmenliğin pek çok boyutuna etki ettiğini söylemektedir (örn. Ball, 1990; Grossman, 1999; Even, 1993; Mason veSpence, 1999; Wilkins, 2008). Pedagojik alan bilgisinin öğretmenliğin pek çok alanına etki yaptığı kabul edilmesine rağmen bu kavramla ilgili pek çok detay henüz net olarak tanımlanamamıştır (Lee veLuft, 2008; Loughran, Mulhall veBerry, 2008).Dolayısı ile öğretmenlik bilgisinin yapısı, alt boyutları, bu boyutların kendi aralarındaki ilişkileri, birbirlerinden nasıl ayrıştığı hala araştırılması gereken konulardandır.

Burada araştırmacıların hemfikir oldukları temel düşüncelerden birisi öğretmenlik mesleğini diğer meslek gruplarından ayıran bir bilgi alanın varlığı ve bu bilgi alanının deneyimle birlikte geliştiğidir. Öğretmenleri diğer meslek gruplarından ayıran bir bilgi alanının varlığında ortak görüşler oluşmuşken, öğretmenlere yönelik hazırlanan testlerin pedagojik alan bilgisi kapsamında ele alınması bir gerekliliktir. Yapılan bu tez çalışmasının sonucunda ortaya çıkacak ürünler kadar, izlenilen yöntemlerin detaylı açıklanması öğretmenlere yönelik geliştirilecek testler ve yapılacak diğer çalışmalar için yol gösterici olacaktır.

Bu çalışmanın ilgili alan yazınına bir diğer katkısı ise özellikle Türkiye'de yapılan çalışmalar arasında kapsam ve araştırma deseni olarak öncü çalışmalardan birisi olmasıdır. Türkiye de özellikle pedagojik alan bilgisini araştıran çalışmaların sayısı yavaş yavaş artmaktadır. Bir diğer özelliği ise Türkiye'de özellikle öğretmenlik bilgisini ölçmeye yönelik geliştirilen ilk test çalışmasıdır. Çalışmanın sonuçlarının program geliştirenler, akademisyenler ve bürokratlar için uygulamaya yönelik sonuçlar ortaya koyması, ve literatürde var olan bu açığa örnek çözümler sunabilmesi hedeflenmiştir.

Bu amaçla, araştırma kapsamında geliştirilecek enstrümanla öğretmen adaylarının ölçme kavramlarını öğretme bilgilerini nicel olarak ölçebilmesi hedeflenmektedir. Bu kapsamda öğretmenlik bilgisi içinde yer alan alt boyutların detaylı olarak incelenmesine olanak verecektir.

BÖLÜM 2

ALANYAZIN ÇALIŞMALARININ İNCELENMESİ

2.1. Öğretmenlik Bilgisi

Öğretmenlik bilgisi konusunda çalışan araştırmacılarortaya koyduğu ortak sonucu şu şekilde özetlemek mümkündür. Araştırmacılar öğretmenlik mesleğinin içinde diğermesleklerdeki gibi (örneğin:mühendislik, tıp doktorluğu) kendine meslek bilgisinin olduğunu ve bunu öğretmenlik bilgisi diye tanımlandığını belirtmişlerdir. (Ball, Lubienski, &Mewborn, 2001; Ball, Hill, &Bass, 2005).Yine bu konudaki araştırmacılar öğretmenlik bilgisinin her disiplin için farklı gereklilikleri olması gerektiğini söylemekle beraber, öğretmenleri konu hakkında bilgi sahibi diğer yetişkinlerden ayırdığını belirterek karakterize etmişlerdir.

Shulman'ın1986 ve 1987 yıllarında yaptığı çalışmalar özellikle öğretmenlik bilgisini tanımlama konusunda çıkış noktası olmuş, sonrasında konuyla ilgili yapılan pekçok çalışmaya yol gösterici olarak ışık tutmuştur. Özetlemek gerekirse,Shulman 1987 yılındaki çalışmasında öğretmenlik bilgisini yedi başlık altında kategorize etmiş ve öğretmenlik bilgisinin bileşenlerini şu şekilde isimlendirmiştir:

1) Alan Bilgisi,

2) Pedagojik Alan Bilgisi,

3) Müfredat Bilgisi,

4) Genel Pedagoji Bilgisi,

5) Öğrencilerin ve Öğrenci Özellikleri Bilgisi,

6) Eğitim Ortamı Bilgisi ve son olarak

7) Eğitim ile ilgili Amaçlar, Değerler, Felsefi ve Tarihsel Süreç Bilgisi.

Shulman yaptığı ilk çalışmasında (1986) öğretmenlerin, mesleki bilgilerini genel olarak: *alan bilgisi, pedagoji bilgisi* ve *müfredat bilgisi* olarak üç temel başlıkta sınıflandırmıştır. Daha sonra ortaya koyduğu modelde bazı düzeltmeler yapmış ve üç bilgi alanının ismini *alan bilgisi, pedagojik alan bilgisi* ve *müfredat bilgisi* olarak değiştirmiştir. Shulman'ın yaptığı bu tanımlamadan sonra Grossman (1990) öğretmenlik bilgisine yeni bir açılım daha getirmiştir. Grossman (1990) ilgili çalışmasında, öğretmenlik bilgisini: *genel pedagoji bilgisi, alan bilgisi, pedagojik alan bilgisi*ve *ortam bilgisi* olarak dört ana başlık altında toplamıştır. Daha sonraki dönemlerde öğretmenlik bilgisini tanımlaya ve sınıflandırmaya yönelik çalışmalarda yukarıdaki örnekte olduğu gibi öğretmenlik meslek bilgisinin kapsam ve isimlendirilmeleri zaman zaman değişime uğramıştır. Fakat pedagojik alan bilgisi her çalışmadaayrı başlık altındayerini almıştır. Her ne kadar içerik ve diğer bilgi alanları ile olan ilişkisi netleşmemiş olsa da bütün çalışmalarda hem fikir olunan nokta pedagojik alan bilgisiningenel pedagoji bilgisi ile alan bilgisinin harmanlanıp ortaya çıktığı yeni bir bilgi formu olarak kabul edilmesidir.Shulman'ın çalışmalarında ortaya koyduğu öğretmenlik bilgisinin bileşenleri ile ilgili genel bilgiler aşağıda başlıklar halinde sunulmuştur.

(a) *Alan Bilgisi*: (Content Knowledge) öğretmenlik yapılan alanla ilgili temel bilimsel bilgileri içeren alan,

(b) *Pedagojik Alan Bilgisi*: (Pedagogical Content Knowledge) alan bilgisi ile pedagoji bilgisi arasında köprü konumunda olan ve bu iki bilgi alanının kaynaşması ile ortaya çıkan yeni bir alan,

(c) *Müfredat Bilgisi*: Öğretmenlik yapılan disiplinle ilgili konu sıralamaları,kavramların birbirleri ile olan ilişkileri, derslerde verilmesi gereken sıralama bilgilerini içeren alan,

(d) *Genel Pedagoji Bilgisi*: Sınıf yönetimi, rehberlik gibi öğrencilerin yaş seviyelerine ve gelişimsel süreçlerine bağlı olarak davranış stratejilerini içeren alan,

(e) Öğrencilerin ve onların özelliklerinin bilgisi,

(f) Eğitim ortamı ile ilgili bilgiler,

(g) Eğitimin genel amaçlarını, daha alt boyuttaki kazanımlarını ve bu amaçların felsefi ve tarihsel temelleri ile ilgili bilgiler.

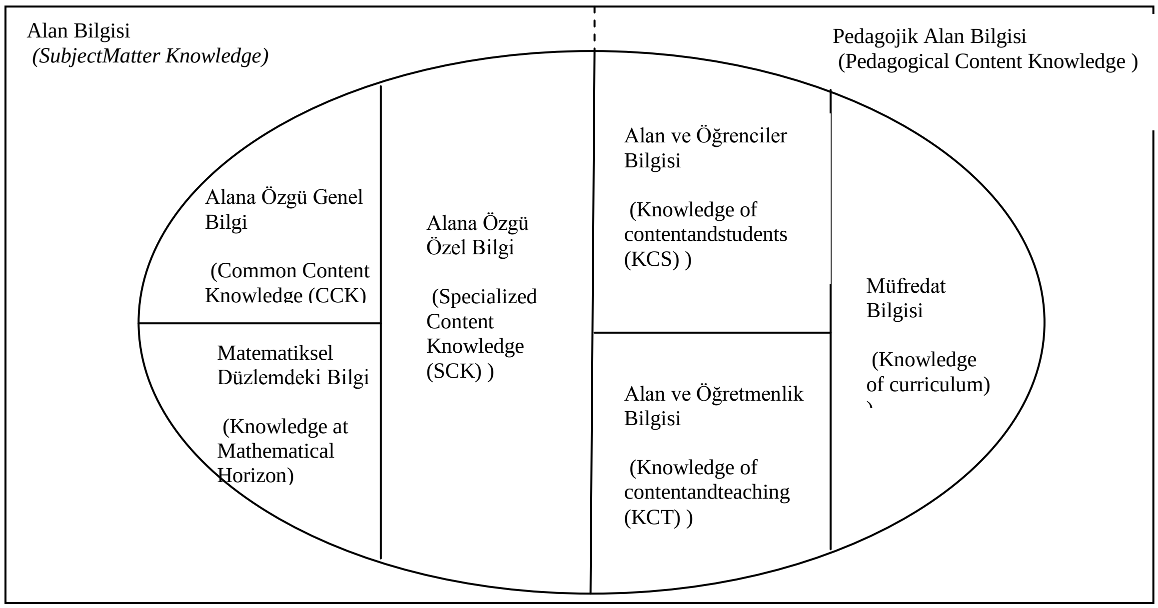
2.2. Pedagojik Alan Bilgisi

Yukarıdaki sınıflandırmada öğretmenlik meslek bilgisi ile doğrudan ilişkili*pedagojik alan bilgisi*niShulman 1986 yılındaki çalışmasında şu şekilde karakterize etmiştir;

> [Pedagojik alan bilgisi] konu ve kavramların en işlevsel gösterimlerini bilme; konuların öğrenilmesini nelerin kolaylaştırdığı yada zorlaştırdığını bilme; öğrencelerin kavram yanılgılarını bilme; kavramların anlaşılması ve kavramsal yanılgıların giderilmesine yönelik analojiler, temsiller, örnekler, açıklamaları bilme; farklı yaştaki ve farklı seviyedeki öğrencilerin kavramlarla ilgili düşünce, algı ve önbilgilerini bilme[dir] (s. 9)

Pedagojik alan bilgisinin hem alan bilgisi hem de genel pedagoji bilgisinden ayıran özelliklerini matematik özelinde şu şekilde örneklemek mümkündür. Bir matematik öğretmenin bir matematik uzmanından farklı olarak ayrıca bilmesi gereken bilgiler, pedagojik alan bilgisinin alan bilgisinden ayrışan tarafını temsil etmektedir. Benzer şekilde pedagojik alan bilgisini, genel pedagoji bilgisinden matematik öğretmenlerini herhangi bir branştaki öğretmenlerden ayıran bilgiler olarak ayrıştırabiliriz.

Hill, Ball, ve Schilling, (2008), Shulman (1986, 1987) yıllarında sunmuş olduğu daha çok teorik yapıdaki *öğretmenlik bilgisini* uygulama yönelik olarak özellikle matematik öğretimi yeniden açarakaşağıdaki modeli sunmuşlardır. Sundukları modelde alan bilgisi ve pedagojik alan bilgisini iki ayrı başlık olarak tanımladıktan sonra her bir bilgi alanını üç ayrıkategoride tanımlamışlardır.



Figür 2.1 Matematik öğretimine yönelik öğretmenlik bilgisi şeması

Ball, ThamesvePhelps, (2008) "Content Knowledge forTeachingWhatMakesIt Special",Journal of TeacherEducation, 59 (5), p. 389–407 makalesinden aynen alınmıştır.

Figür 2.1'de verilen modeldeki kavramlardan kısaca bahsedilmesi gerekirse;*Matematiksel Düzlemdeki Bilgi*, çok temel matematiksel bilgilerini içerir ve öğretmenlik bilgisi ile doğrudan ilgili değildir.

Alana Özgü Genel Bilgi: Shulman'ın tanımladığı *Alan Bilgisi*ile paralellik göstermektedir. Bir matematik öğretmeninin genel olarak matematiksel kavramlarla ilgili sahip olması gereken bilimsel bilgileri içermektedir.

Alana Özgü Özel Bilgi: Matematik öğretmenlerinin öğretecekleri kavramlarla ilgili sahip olması gereken bilimsel bilgileri içermektedir. Bu bilgi alanı yeni tanımlananbir kavram olmakla birlikte, sınıf ortamında ele alınması gerekenmatematiksel kavramlar çerçevesinde tanımlanmıştır. Sınıfta işlenen her hangi bir matematiksel gösterimin temsil etmesi gereken matematiksel düşünceyi ne kadar yansıttığını, matematiksel düşünceyi temsil ederken ne oranda zayıf kaldığını anlamaya yarayan bilgileri içermektedir. Aynı zamanda sınıf ortamındakarşılaşılanfarklı matematiksel çözüm yöntemlerinin matematiksel doğruluğunu ve genellenebilirliğinitest edebilme bilgilerini içermektedir (Ball, Hill&Bass, 2005, p 378).

Alan ve Öğrenciler Bilgisi: Matematik öğretmenlerinin öğrencilerin yaş seviyelerine uygun olarak onların karakteristiklerive gelişimsel süreçleri ile ilgili bilgileri kapsamaktadır. Öğrenci karakteristiklerinin yanı sıra onların hazırbulunuşluk düzeyleri, öğrencilerin konuya özgü olarak algı şekilleri, bir konuyu anlamada yaşayabilecekleri zorluklar ve kavram yanılgıları ile ilgili sahip olması gereken bilgileri içermektedir.

Alan ve Öğretmenlik Bilgisi: Matematik öğretmenlerinin işleyecekleri derslerde konu ve kavramların en işlevsel gösterimlerini, konuları işlerken kavramların anlaşılmasını kolaylaştıran ya da zorlaştıran etkenleri, kavram öğretiminde anlatımı kolaylaştıran analojiler, örnekler, gösterimleri açıklamalar, kavram yanılgılarının giderilmesine yönelik kullanılabilecek örnekleri ya da karşıt örnekler ile ilgili sahip olması gereken bilgileri içermektedir.

Müfredat Bilgisi: Matematiksel kavramların sıralamaları, matematiksel kavramların birbirleri ile olan ilişkilendirmeleri, disiplin içi ya da disiplinler arası ilişkilendirmeler ile ilgili bilgileri içermektedir.

2.3. Pedagojik Alan Bilgisinin Ölçülmesi

Hill, Sleep, Lewis, ve Ball (2007) çalışmalarında öğretmenlik bilgisinin geçerli ve güvenilir şekilde ölçülebilmesinin gerekçelerini üç temel başlık altında toplamaktadırlar. İlki bürokratik bir gerekçe olarak, öğretmen atamaları sırasında daha donanımlı, mesleğini gerektirdiği becerilere sahip öğretmenleri geçerli ve güvenilir şekilde atayabilmek bürokratların çok uzun yıllardır çözüm aradığı sorunlardan birisidir. Bu anlamda öğretmen atamalarında kullanılabilecek geçerli güvenilir ölçme araçlarına ihtiyaç vardır. İkinci olarak eğitim fakültelerinde verilen eğitimin yeterli olup olmadığını test etmek, öğretmen adaylarının öğretmenlik mesleği için gerekli becerilere sahip olup olmadıklarını değerlendirmek, eğitim fakültelerinin etkililiğini test etmek için akademisyenler öğretmenlik bilgisini ölçebilecek ölçme araçlarına ihtiyaç duymaktadırlar. Hill ve arkadaşlarının (2007) çalışmasında Pedagojik alan bilgisinin ölçülmesine yönelik bahsi geçen bir diğer gerekçe ise yine akademik amaçlıdır. Öğretmenlik bilgisininyapısınıanlayabilmek, hem kendi içindehem de diğer bilimsel yapılarla (construct) olan ilişkilerini ortaya çıkarabilmesiiçin bu kavramın ölçülmesine ve dolayısı ile ölçme araçlarına ihtiyaç duyulmaktadır. Teorisyenler kavramın ölçülebilmesine bağlı olarak öne sürülen teorilerin geçerliliğini test edebileceklerini, ölçme sonuçlarına bağlı olarak öne sürülen teorileri iyileştirebileceklerini düşünmektedirler.

Fakat yine aynı çalışmada belirtildiğine göre günümüze kadar öğretmenlik bilgisini ölçmeye yönelikyapılan çalışmaların öğretmenlik bilgisini ölçmeye yönelik olarak amacına hizmet etmediği, elde edilen sonuçların öğretmenlik bilgisi açısından problemli olduğunu söylemektedir. Bu kapsamda öğretmenlik bilgisini test etmek amacı ile öğretmenlerin lisans dönemlerinde aldıkları alan bilgisi ders sayısı, derslerde aldıkları notlar, sadece matematikalan bilgisine yöneliksınav sonuçları ya da tam tersi sadece pedagojik bilgilerini yönelik hazırlanmış sınavlar sonuçlarının öğretmenlik bilgisine yönelik çok anlamlı sonuçlar vermediği gözlemlenmiştir (Monk, 1994)

Özellikle pedagojik alan bilgisi tanımının yapılmasından sonra öğretmenlik bilgisi ile ilgili çalışmalar daha çok pedagojik alan bilgisi üzerine odaklanmıştır. Pedagojik alan bilgisini ölçmeye yönelik eğitim alanında belli bir birikim oluşmaya başlamıştır. Bu anlamda literatürde pedagojik alan bilgisini ölçmeye yönelik yapılan çalışmaları genel olarak üç başlık altında sınıflandırmak mümkündür.

İlki kağıt kalem testleridir, bu kategoriye Likert türü tutum ölçekleri, çoktan seçmeli testler, açık uçlu ve diğer soru tiplerinde hazırlanmış yazılı sınavların hepsi girmektedir. Bu alandaki soruların hepsi öğretmenlerin tutumları, inançları ve konu ile ilgili alan bilgilerini ölçmeye yöneliktir.. Kalem kağıt testlerinin uygulaması ve çok sayıda öğretmene ulaşabilmesi gibi avantajları olmasına rağmen, hazırlanan ölçeklerdeyapı geçerliliğini sağlamanın ve güvenirlik değerlerini yüksek tutmanın zor olduğunu söylenmektedir. Bir diğer husus ise, özellikle çoktan seçmeli olarak hazırlanan test soruları ile öğretmenlik gibi kendi içinde gayet karmaşık ve çok değişkeni olan bir kavramı işin özünü kaçırmadan ölçülebilmesinin bir hayli zor olmasıdır (Hill ve diğer, 2008).

Pedagojik alan bilgisini ölçmeye yönelik tasarlanan bir diğer kategori ise, kavram haritaları, resimler, grafiklerle yapılan çalışmalardır. Bu kategori öğretmenlerin bilgilerini nasıl ilişkilendirdiklerini ortaya çıkarmaya yönelik çalışmaları içermektedir. Aynı kavramlar üzerinde farklı öğretmelerle yapılan çalışmalarda öğretmenlerin beyinlerinde oluşturdukları ilişkilendirmeleri görsel hale getirmeye çalışmaktadır (Baxter&Lederman, 1999). Fakat Kagan (1990) bu yöntemin kısa soluklu çalışmalar için kullanıldığını ve ortaya çıkan sonuçların öğretmenlerde kısa dönemlideğişiklikleri ortaya çıkarmaya yönelik olduğunu belirtmektedir. Bu kategorideki çalışmalarla ilgili bir diğer eleştiri ise, öğretmenlerin çalışmalar sırasında ortaya koydukları ürünlerin sahip oldukları gerçek düşünceler olup olmadığını test etmenin oldukça zor olduğudur.

Pedagojik alan bilgisini ölçmek amacı ile kullanılan son ölçme kategorisi ise içerisinde birden fazla ölçme değerlendirme metodunu barındıran karma metotlardır. Bu kategori, yazılı sınavlardan, mülakata, gözlemlerden görüşmelere kadar bütün halde uygulanan uzun soluklu çalışmaları içermektedir (Baxter&Lederman, 1999). Bahsedilen bu son kategorinin içerik olarak zengin ve oldukça geçerli sonuçlar verse de, uygulamada öğretmen değerlendirmesi açısından hantal bir yapısı olduğu açıktır. Hatta bürokratik olarak çok sayıda öğretmenin işe alım sürecinde uygulanması neredeyse imkânsızdır. Ayrıca bu çalışmaların replike çalışmalarını yapmak oldukça zordur (Baxter&Lederman, 1999).

BÖLÜM 3

ÇALIŞMANIN YÖNTEMİ

Bu çalışmanın esas amacı ilköğretim matematik öğretmen adaylarının ölçme kavramlarını özellikle uzunluk, alan ve hacim ölçme kavramlarını öğretme bilgilerini ölçmeye yönelik çok seçmeli bir test geliştirmekti. Bu amaçla Downing (2006) çalışmasında özetlemiş olduğu test geliştirme basamakları birebir takip edildi. Test geliştirme çalışması yapılırken, 2010 yılı Sonbahar döneminden 2012 yılı bahar dönemine kadar veri toplanması ve bu verilen tekrarlı olarak analiz edilmesi hedeflendi. Çalışma kapsamında 4 ana aşamada veri toplandı. 1. aşamada madde geliştirme ve pilot çalışmalar yapıldı. 2. ve 3. aşamalarda maddelerin alan uygulamaları ve çıkan sonuçlara göre revizyonları yapıldı. 4. aşamada ise çalışmanın geçerlilik analizleri yapıldı. Çalışmanın katılımcıları Türkiye'nin farklı bölgelerindeki 20 ilköğretim matematik öğretmenliği bölümlerindeki 4. sınıf öğrencileri ile çalışıldı. Veri toplama süreci Tablo 3.1'de özetlenmiştir. Kısaca bu şekilde özetlenen çalışma yöntemi ile ilgili bilgiler bu başlık altında açıklanacaktır.

Table 3.1 Test uygulama sürecinin özeti

Aşama	*Zaman Aralığı*	*Katılımcılar*	*Amaç*	*Enstrüman*	*Veri Türü*	*Analiz*
Uygulama I	Eylül 2010- Aralık 2010	44 Öğretmen Adayı	Test Geliştirme ve	Testin 29 maddelik versiyonu	Nitel ve Nicel Veri	Tema Analizi
		27 Öğretmen Adayı	Pilot Uygulama	Yarı yapılandırılmış görüşme		
Uygulama II	Şubat 2011- Mayıs 2011	1010 Öğretmen Adayı	Revizyon	Testin 16 maddelik versiyonu – iki ayrı test formu	Nicel Veri	Madde Analizi ve Rasch Analizi
Uygulama III	Eylül 2011- Ekim 2011	99 Öğretmen Adayı	Revizyon	Testin 20 maddelik versiyonu	Nicel Veri	Madde Analizi ve Rasch Analizi
Uygulama IV	Şubat 2012- Nisan 2012	167 Öğretmen Adayı	Geçerlilik çalışması	Testin 15 maddelik versiyonu	Nicel Veri	Madde Analizi ve Rasch Analizi

3.1. Uygulama I

Çalışmanın birinci aşamasındaki amaç geliştirilen maddelerin nasıl çalıştıklarını görmek soru maddelerinin hedeflenen yapısı ile örtüşüp örtüşmediğini test etmekti. Bu amaçla test maddeleri hakkında detaylı bilgiler elde etmek için 2010 yılı Güz döneminde geliştirilen test maddelerine dair nitel ve nicel iki tür veri toplanmıştır.

İlk olarak bir devlet üniversitesindeki 44 (32 Kadın, 12 Erkek) öğretmen adayına 29 tane çoktan seçmeli madde yöneltilmiştir.Onlardan çoktan seçmeli maddelere yanıt vermelerinin yanı sıra her bir soru için tasarlanan açık uçlu bölümlerde (Bkz Ekler A), her bir soru ile ilgili nasıl düşündüklerini, nasıl cevaplama stratejileri geliştirdiklerini ve soru içinde kurgulanan problemle ilgili düşüncelerini detaylı şekilde yazmaları istenmiştir.

Bu uygulamadan sonra 27 (17 Kadın, 10 Erkek) öğretmen adayı ile yarı yapılandırılmış 40-60 dk süren görüşmeler düzenlemiştir. Bu görüşmelerin amacı, öğretmen adaylarının sorular karşısında nasıl bir düşünce geliştirdiklerini yakından görmek ve geliştirilen test maddelerinin tasarlandığı şekilde çalışıp çalışmadığını detaylı inceleme fırsatı yakalamaktı.

Toplanan veriler hem nicel hem de nitel analiz yöntemleri ile incelenmiş, elden edilen bulgular ışığında test maddeleri revize edilmiştir. Tam çalışmadığı gözlemlenen 1 madde elenmiş, benzer yapıdaki maddeler birleştirilerek toplam 26 maddelik yeni bir madde seti oluşturulmuştur.

3.2. Uygulama II

İkinci aşamadaki temel amaç geliştirilen test maddelerinin Klasik Test Teorisine ve Madde Tepki Kuramına göre nasıl çalıştığını incelemekti. Bu aşamada Uygulama I'den elde edilen bulgular ışığında revize edişmiş 26 maddelik yeni madde seti uygulanmıştır. Fakat Uygulama I'dekiönemli gözlemlerden bir tanesi uygulanan testin bir ders saatinden fazla zaman gerektiriyor olmasıdır.Bu problemi aşmak adına 26 soruluk yeni madde setinin 6 maddesi ortak olacak şekilde, iki ayrı paralel set halinde uygulanılmasına karar verilmiştir.

Bu aşamada özellikle Madde tepki Kuramı açısından anlamlı sonuçlar elde edebilmek için mümkün mertebe çok sayıda öğretmen adayına ulaşmak hedeflendi. Bu amaçla, Türkiye'nin hemen her bölgesinden olacak şekilde bilgisayar ortamında rasgele seçilmiş 17 üniversitelerdeki ilköğretim öğretmenliği fakültesinden 1010 öğretmen adayı ile çalışılmıştır. Katılımcılara ait bilgilerin özeti Tablo 3.1'de sunulmuştur.

Tablo 3.1Katılımcıların demografik bilgilerinin frekans dağılımı (n=1010)

	Frekans (f)	Yüzde (%)
Cinsiyet		
Erkek	346	34.3
Kadın	649	64.3
Kayıp Değer	15	1.5
Toplam	1010	100
Program Cinsi		
1. Öğretim	615	60.9
2. Öğretim	393	38.9
Kayıp Değer	2	0.5
Toplam	1010	100
Öğretmenlik Deneyimi (saat)		
Yok	305	30.2
10 saatten az	115	11.4
11-20 saat	130	12.9
21-50 saat	187	18.5
51-100 saat	114	11.3
101 saatten fazla	159	15.7
Toplam	1010	100
Mezun olduğu lise		
Düz Lise	131	13.0
Anadolu lisesi	252	25.0
Anadolu Öğretmen Lisesi	478	47.3
Yabancı Dil Ağırlıklı Düz Lise	93	9.2
Kayıp Değer	56	5.5
Toplam	1010	100

Elde edilen verilerle hem Klasik Test teorisine göre hem de Madde Tepki kuramına göre madde analizi yapılmıştır. Klasik Test Teorisine göre madde istatistiklerini elde etmek için ITEMAN (Version 3.6 byAssessmentSystems Corporation 1994), Rasch Analizi içinse BIGSTEPS (Wright &Linacre, 1991) paket programları kullanılmıştır.

3.3. Uygulama III

2.Aşamadaki madde analizi sonuçlarına göre 26 madde içinde daha iyi çalışan maddelerle yeni bir soru seti oluşturuldu. Oluşturulan 20 soruluk bu yeni setteki maddelerin bütüncül olarak nasıl çalıştığını görmek amacı ile Ankara'daki 2 üniversitede bulunan ilköğretim matematik öğretmenliği fakültesi 4. Sınıf öğrencileri ile 3. Aşama uygulaması yapıldı. Bu aşamada çalışmaya 99 öğretmen adayı (79 Kadın, 19 Erkek, 1 Kayıp değer) katıldı.İkinci aşamadaki veri analizine benzer şekilde bu aşamada da Klasik Test Teorisi ve Madde Tepki kuramına göre madde analizleri yapıldı.Veri analizi sonuçlarına dayanarak, en iyi çalışan 15 madde ile yeni bir set oluşturuldu ve bu set geçerlilik çalışmaları için 4. Aşamada kullanıldı.

3.4. Uygulama IV

Son ve 4. Aşamadaki oluşturulan son madde seti 2011- 2012 öğretim yılı Bahar döneminde Ankara'da bulunan 3 üniversitedeki (1 Özel, 2 Devlet) ilköğretim matematik öğretmenliği bölümlerindeki 168 (146 Kadın, 22 Erkek) öğretmen adayı ile gerçekleştirildi. 2. ve 3. Aşamadaki uygulamalardan farklı olarak, bu aşamada özel öğretim yöntemleri derslerini tamamlarmış 3. Sınıf öğrenciler de çalışmaya dahil edildi. Soru setinin 15 maddelik son hali öğretmen adaylarına sunuldu, ve elde edilen veriler önceki uygulamalara paralel olarak Klasik Test Teorisi ve Madde Tepki kuramında göre analiz edildi.

BÖLÜM 4

SONUÇLAR

Test geliştirme süreci sürekli veri toplamayı, toplanan verileri analiz etmeyi ve sonuçlar ışığında yeni değişiklikler yapıp yeni uygulamalar yapmayı gerektiren bir süreçti. Bu amaçla, 4 aşamada toplanan veriler ve analiz sonuçları bu bölümde kısaca özetlenecektir.

4.1. Uygulama I

Maddelerin Geliştirilmesi ve Pilot Çalışması

Birinci aşamadaki temel amaç geliştirilen maddelerin istendik şekilde çalışıp çalışmadığını test etmekti. Bu amaçla 29 maddelik soru setini kullanarak nitel ve nicel iki farklı veri toplandı. Nicel olarak madde istatistikleri belirlenirken, nitel data analizi kapsamında açık uçlu sorulara verilen cevapların içerik analizleri ve yarı yapılandırılmış görüşmelerin nitel analizleri yapılmıştır. Bu analizlerin sonuçları, maddelerin revizyonlarında ve geniş kapsamlı uygulanacak yeni soru madde setlerin oluşturulmasında kullanılmıştır.

Açık uçlu sorulara öğretmen adaylarının vermiş olduğu cevaplar 0-2 aralığına sahip derecelendirilmiş puanlama anahtarı aracılığı ile kodlanmış. Burada 0, tamamen yanlış, yada boş cevaplar için kullanılırken; 2 puan içinde senaryolaştırılmış problem durumunun tam tespiti, matematiksel olarak doğru müdahale yöntemlerini değerlendirmek için kullanılmıştır. Öğretmen adaylarının çoktan seçmeli sorulara vermiş oldukları yanıtlar ile yaptıkları açıklamaların ortalamalarını gösteren sonuçlar Tablo 4.1'de sunulmuştur. Bu tablodaki değerlere göre yanlış cevaplanmış fakat kısmen doğru ya da tam doğru gerekçelendirilmiş maddeler bir kez daha elden geçirilmiştir. Doğru- yanlış sayısı, açık uçlu cevapların ortalamaları ve madde üzerinde yapılan değişiklik Tablo 4.1'de verilmiştir.

Tablo 4.1 Açık uçlu sorulara verilen cevapların içerik analizi ve bağlı değişiklikler

	Doğru		*Yanlış*		
Madde	*#*	*Ortalama*	*#*	*Ortalama*	*Değişiklik*
VP1		1.	2	1.	Maddenin yapısı değişti
	2	14		04	(KMY*)
LS2	19	1.53		0.64	Soru kalıbı ve görsel değişti.
AP3	34	1.26		0.70	Değişiklik yok.
AP4	5	1.46		1.15	Maddenin yapısı değişti (KMY*)
AP5	32	0.69		0.50	Soru kalıbı ve görsel değişti.
LS6	8	1.38		0.14	Soru kalıbı değişti.
LAP7	27	1.52		1.10	Çeldiriciler LP28 ile birleştirildi.
MS8	13	1.10		0.74	Çeldiriciler değişti.
VS9	27	0.61		0.24	Soru kalıbı vurgulu yapıldı.
AP10	7	1.43		0.65	Soru kalıbı ve çeldiriciler değişti.
LS11	28	1.29		0.93	Çeldiriciler ve görsel değişti.
LS12	28	0.82		0.56	Çeldiriciler ve görsel değişti.
AS13	18	0.56		0.38	Soru kalıbı vurgulu yapıldı.
LAP14	10	1.5		0.38	Soru kalıbı değişti.
AP15	38	1.47		0.00	Görsel değişti.
LS16	14	1.5		0.24	Değişiklik yok.
AP17	0	0.00		0.74	Değişiklik yok.
VS18	29	1.72		0.67	Değişiklik yok.
AS19	23	1.00		0.29	Soru kalıbı ve görsel değişti
MS20	17	1.12		0.52	Soru kalıbı ve çeldiriciler değişti.
AP21	7	1.29		0.62	Maddenin yapısı değişti (BMY**)
AP22		Elendi			
AS23	38	0.82		0.33	Değişiklik yok.
VS24	15	0.93		0.10	Soru kalıbı değişti.
LAS25	37	1.05		0.29	Görsel değişti.
AVP26	26	1.07		0.33	Çeldiriciler ve görsel değişti.
VS27	11	0.82		0.67	Çeldiriciler değişti.
LP28	31	1.48		0.46	Çeldiriciler LAP7 ile birleştirildi.
LP29	20	1.35		0.58	Görsel değişti.

KMY*: Kompleks çoktan seçmeli madde yapısı, BMY**: Basit çoktan seçmeli madde yapısı

4.2. Uygulama II

Sonuçlar

İkinci uygulamanın temel amacı geliştirilen maddelerin Klasik Test Teorisi ve Madde Tepki Kuramı çerçevesinden nasıl çalıştığını ortaya çıkarmaktı. Bu sebeple 6'sı ortak olmak üzere 16 maddeden oluşan iki ayrı setleri için Türkiye genelindeki 17 üniversiteden toplanan veri ile analizleri yapıldı.Hem Klasik Test Teorisi hem de RaschAnalizi'nde kullanılan araçlar kullanılarak her bir madde için istatistikler hesaplandı.

4.2.1. Test 1 ve Test 2 için Tek Boyutluluk Analiz Sonuçları

Rasch Analizi kullanılarak Test 1 ve Test 2 içindeki maddelerin tek bir yapısal boyutta olup olmadıkları test edildi. Analiz sonuçları hem Test 1 içindeki 16 madde hem de Test 2 içindeki 16 maddenin fit değerlerinin beklenen 0.7 - 1.3 (Linacre, 2007) aralığında olduğu gözlemlenmiştir. Maddelerin ilgili bilgilerini özeti Figür 4.1 ve Figür 4.2'de sunulmuştur.

```
ITEMS FIT GRAPH:  MISFIT ORDER

+--------------------------------------------------------------------+
|ENTRY| MEASURE |  INFIT MEAN-SQUARE  | OUTFIT MEAN-SQUARE  |        |
|NUMBR| -     + |0     0.7 1 1.3     2|0     0.7 1 1.3     2| ITEMS  |
|-----+---------+---------------------+---------------------+--------|
|    9 |      *  |       :  *  :       |A      :  *  :       | AP5    |
|   12|     *   |       :  *  :       |B      :  *  :       | VS27   |
|   15|      *  |       : *.  :       |C      :  *  :       | LS12   |
|   11|   *     |       :  *  :       |D      :  *  :       | MS8    |
|   13|   *     |       :  *  :       |E      :  *  :       | VS18   |
|    2|    *    |       :  *  :       |F      :  *  :       | AP3    |
|   10|        *|       :  *  :       |G      :  *  :       | AVP26  |
|   16|    *    |       : *.  :       |H      :  *  :       | VS24   |
|    5|   *     |       :  *  :       |h      : *.  :       | AS19   |
|    8|     *   |       :  *  :       |g      : *.  :       | AP17   |
|    6|       * |       : *.  :       |f      :* .  :       | LAP14  |
|    7|      *  |       : *.  :       |e      : *.  :       | MS20   |
|    4|     *   |       : *.  :       |d      : *.  :       | AP4    |
|    3|     *   |       : *.  :       |c      : *.  :       | AP15   |
|   14|     *   |       : *.  :       |b      : *.  :       | AP10   |
|    1 |*        |       : *.  :       |a      :* .  :       | LS2    |
+--------------------------------------------------------------------+
```

Figür 4.1 Maddelerin tek boyutluluk göstergesi (Test 1)

ITEMS STATISTICS: MISFIT ORDER

ENTRY NUMBR	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBIS CORR.	ITEMS
14	379	395	-4.20	.25	1.13	.5	.93	-.2	A .03	LAS25
9	119	383	-.07	.11	1.05	1.0	1.05	.6	B-.07	AVP26
16	127	389	-.15	.11	1.03	.7	1.01	.1	C-.05	AP4
3	80	398	.51	.12	1.03	.4	1.02	.1	D-.04	AP13
11	150	395	-.38	.10	1.02	.5	1.01	.2	E-.02	LAP7
7	95	395	.28	.11	1.01	.2	.98	-.2	F-.02	AP21
10	50	396	1.06	.15	1.00	.0	.90	-.6	G-.01	VS27
8	172	396	-.63	.10	1.00	.2	.97	-.5	H .00	MS20
15	88	356	.23	.12	1.00	-.1	.96	-.4	h .01	LS16
5	64	396	.78	.13	.99	-.1	.94	-.4	g .02	LAP14
6	37	389	1.37	.16	.99	-.1	.83	-.9	f .05	LS11
1	53	394	.99	.14	.98	-.1	.91	-.6	e .03	VP1
4	67	301	.37	.13	.98	-.2	.95	-.4	d .04	VS9
13	122	393	-.07	.11	.98	-.4	.98	-.3	c .02	MS8
12	133	393	-.21	.10	.97	-.6	.94	-.9	b .04	LP29
2	99	363	.12	.11	.97	-.6	.94	-.7	a .05	LS6
MEAN	115.	383.	.00	.13	1.01	.1	.96	-.3		
S.D.	78.	24.	1.21	.04	.04	.4	.05	.4		

Figür 4.2 Maddelerin tek boyutluluk göstergesi (Test 2)

4.2.2. Test 1 ve Test 2 için Madde-Kişi Grafiği

Rasch Analizinin kullanışlı göstergelerinden birisi de maddeler ve kişilerin yetenek kestirimlerini tek bir grafik üzerinde özetleyebilmesidir. Bu grafik bize maddelerin kişilerle nasıl eşleştiğini, madde zorluğunu ve kişilerin yetenek kestirimlerinin sorular bazındaki dağılımı vermektedir. Bu bilgilerin özetleri Figür 4.3 ve Figür 4.4 özetlenmektedir.

```
         MAP OF PERSONS AND ITEMS
               <frequ>  |  <less>
       3                +
                        |
                        |
                        |
                        |
                        |
                        |Q AP17
                        |
       2                +
                        |
                        |
                        |
                        |
                        |  AP4
                        |
                        |S
       1             .  +  AP10
                        |  VS27
                        |
                    .#  |  VS18
                     . Q|  LAP!4
                    .#  |
                     .  |  AS19
                     .  |  LS2
       0           .##  +M AP5
                     .  |  MS8
              .#######  |  AVP26
                     . S|
                     .  |
               .######  |  VS24
                    .#  |  MS20
                     .  |
      -1  .###########  +  LS12
                       M|S
                    .#  |
          .###########  |  AP3
                    .#  |
                        |
                   .##  |
              .####### S|
      -2                +
                     .  |
                        |Q
                 .####  |
                        |
                       Q|
                        |
                        |  AP15
      -3                +
                     .  |
                        |
                     #  |
                        |
                        |
                        |
                        |
      -4             .  +
                  <rare>|<more>

EACH '#' IN THE PERSON COLUMN IS   3 PERSONS; EACH '.' IS 1 TO   2 PERSONS
```

Figür 4.3 Soru ve kişi dağılım göstergesi (Test 1)

Figür 4.3’te görüldüğü üzere maddeler (-3, +3) aralığında dengeli bir dağılım göstermiştir. Bir yerde kümelenme olmaması her bir maddenin farklı bir zorluk derecesine sahip olduğunu göstermektedir. Aynı figür üzerinde en kolay maddenin AP15, en zor maddenin ise AP17 olduğu kolaylıklar gözlenebilir. Öte yandan, kişilerin test sonuçlarının dağılımı ise (-4, +1) aralığında olup beklenen (-3,+3) aralığından sapma göstermiştir. Bu sonuçlara göre AP4, AP17 maddeleri katılımcıların yeteneklerinin üzerinde kalmıştır.

Test 1’in sonuçlarına benzer şekilde Test 2 içindeki maddeler de benzer bir dağılım göstermiştir. Test 2 içindeki maddelerin güçlük değerleri yüksek çıkmış olsa da lineer gösterge üzerinde beklendik aralıklar içinde dengeli bir dağılıma sahiptir. Bu set içindeki maddeler (-3,+2)) aralığında dağılım gösterirken, bu soru setini cevaplayan katılımcıların yetenek kestirimleri (-4, +1) aralığında olmuştur. Bu set içindeki yukarda olan LS11 en zor madde olarak gözlemlenirken, en altta yer alan LAS 25 maddesi ise en kolay madde olmuştur.

```
               MAP OF PERSONS AND ITEMS
                            <more>  |  <rare>
                   2                +
                                    |
                                    |T
                                    |
                                    |
                                    |
                                    |
                                    |  LS11
                                    |
                                    |
                   1                +
                                    |S VS27
                                    |  VP1  LAP14
                                    |
                                 .  |
                                 .  |  AP13
                                 . T|
                                    |  VS9
                                 .  |  AP21
                                .#  |  LS16
                   0             .  +M LS6
                                 .  |  AVP26
                              .###  |  MS8  AP4
                                 #  |  LP29
                               .## S|
                        .#########  |  LAP7
                                .#  |
                                .#  |  MS20
                      .###########  |
                                 #  |S
                  -1            .#  +
                              #### M|
                      .###########  |
                                 .  |
                                 .  |
                                .#  |
                              .###  |
                          .#######  |
                                 . S|T
                                 .  |
                  -2             .  +
                                 .  |
                                 .  |
                             .####  |
                                    |
                                   T|
                                    |
                                    |
                                    |
                                    |  LAS25
                  -3                +
                                 .  |
                                ##  |
                                    |
                                    |
                                    |
                                    |
                                    |
                  -4             .  +
                          <rare>  |   <more>
EACH '#' IN THE PERSON COLUMN IS   6 PERSONS; EACH '.' IS 1 TO   5 PERSONS
```

Figür 4.4 Soru ve kişi dağılım göstergesi (Test 2)

4.2.3. Test 1 ve Test 2 için Güvenirlik ve Ayırt Edicilik

Figür 4.5 üzerinde özetlenen bilgilerle yola çıkarak Rasch Analizi sonucunda Test 1'e ait madde güvenirlik indeksinin .99 (≥.90) olduğu, maddelerin zorluk olarak yaklaşık 9 kategoriye (with a Separation of 9.12) ayrıldığı görülebilir.

```
SUMMARY OF    16 MEASURED ITEMS
+-----------------------------------------------------------------------------+
|          RAW                                MODEL        INFIT      OUTFIT   |
|         SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN     140.3     462.1         .00     .12      1.00     .1    .97    -.1 |
| S.D.      92.3      25.4        1.14     .03       .03     .6    .09     .8 |
| MAX.     400.0     473.0        2.25     .23      1.06    1.1   1.06    1.4 |
| MIN.      18.0     373.0       -2.82     .09       .95   -1.0    .77   -1.4 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .12  ADJ.SD    1.13  SEPARATION  9.12   ITEM  RELIABILITY  .99 |
|MODEL RMSE    .12  ADJ.SD    1.13  SEPARATION  9.17   ITEM  RELIABILITY  .99 |
| S.E. OF  ITEM  MEAN    .29
+-----------------------------------------------------------------------------+
```

Figür 4.5 Madde Bilgileri Özeti (Test 1)

Benzer şekilde Test 2 içindeki maddeler için güvenirlik indeksinin .98 (≥.90),maddelerin zorluk olarak yaklaşık 6 kategoriye (with a Separation of 6.97) ayrıldığı görülebilir. Bu bilgiler Figür 4.6'da özetlenmiştir.

```
SUMMARY OF     16 MEASURED  ITEMS
+-----------------------------------------------------------------------------+
|          RAW                                MODEL        INFIT      OUTFIT   |
|         SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN     136.3     482.6         .00     .11      1.00    -.1   1.01    -.1 |
| S.D.      76.7      30.0         .76     .01       .02     .5    .08     .9 |
| MAX.     392.0     501.0         .96     .13      1.04     .7   1.14    1.6 |
| MIN.      60.0     380.0       -2.41     .09       .96    -.9    .88   -1.5 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .11  ADJ.SD     .75  SEPARATION  6.97   ITEM  RELIABILITY  .98 |
|MODEL RMSE    .11  ADJ.SD     .75  SEPARATION  7.01   ITEM  RELIABILITY  .98 |
| S.E. OF  ITEM MEAN = .20                                                    |
+-----------------------------------------------------------------------------+
UMEAN=.000 USCALE=1.000
```

Figür 4.6 Madde Bilgileri Özeti (Test 2)

4.2.4. Klasik Test Teorisine göre Test 1 ve Test 2 için Madde İstatistikleri

Klasik Test Teorisinin araçları kullanılarak yapılan analiz sonucunda Uygulama II'denelde edilen bulgular Tablo 4.2 ve Tablo 4.3'te özetlenmiştir.

Madde güçlüğü

ITEMAN sonuçlarındaki (p) değerlerini dikkate aldığımızda maddelerin güçlük olarak (p= .05) değerinden (p=.76) değerine kadar farklı değerlerde olduğu gözlemlenmiştir. Güçlük değeri .40'tan daha düşük maddeler zor maddeler olarak tanımlanmıştır (Haladyna, 2004) ve bu bilgiye göre testin geneli katılımcılar için zor maddeler olmuştur. İdeal güçlük değeri olan .50 civarında üç madde vardır: AP3 (p= .55), MS20 (p= .44), and LS12 (p= .51). Madde güçlüklerinin ortalaması .30 olması Test 1'in genel olarak zor bir test olarak algılandığını göstermektedir.

Tablo 4.2 Madde istatistikleri (Test 1) (n=502)

Madde İstatistikleri				Çeldirici İstatistikleri (%)					
Madde	*p*	*D*	*r*	A	B	C	D	E	Omit
LS2	0.25	0.40	0.52	0.25+	0.18	0.18	0.35	0.03	0.01
AP3	0.55	0.38	0.40	0.27	0.03	0.02	0.55+	0.13	0.00
AP15	0.76	0.41	0.55	0.02	0.02	0.06	0.04	0.76+	0.10
AP4	0.10	0.16	0.41	0.13	0.30	0.10+	0.02	0.45	0.00
AS19	0.23	0.32	0.45	0.15	0.49	0.03	0.09	0.23+	0.01
LAP14	0.19	0.23	0.37	0.16	0.18	0.22	0.19+	0.24	0.01
MS20	0.44	0.41	0.41	0.44+	0.09	0.09	0.17	0.20	0.01
AP17	0.05	0.07	0.34	0.28	0.08	0.37	0.05+	0.20	0.01
AP5	0.28	0.31	0.34	0.28+	0.07	0.02	0.11	0.52	0.00
AVP26	0.27	0.35	0.48	0.33	0.27+	0.22	0.10	0.06	0.02
MS8	0.33	0.32	0.37	0.04	0.39	0.17	0.33+	0.06	0.00
VS27	0.12	0.11	0.24	0.40	0.29	0.12+	0.08	0.10	0.01
VS18	0.15	0.24	0.48	0.22	0.10	0.36	0.15+	0.15	0.02
AP10	0.14	0.22	0.39	0.47	0.14+	0.13	0.21	0.04	0.00
LS12	0.51	0.39	0.38	0.18	0.14	0.04	0.09	0.51+	0.04
VS24	0.36	0.41	0.46	0.14	0.14	0.20	0.36+	0.05	0.11

+: Anahtar

Ayırtedicilik

İki madde ayırtedicilik indeksi bu aşamada maddelerin nasıl çalıştığını belirlemek noktasında kriter olarak değerlendirilmiştir. İlki, madde analizi sonuçlarındaki D değerleri, bir diğer ise point-biserial korelasyonudur. Bu noktada .40 ve üzerindeki D değerlerine sahip maddeler ayırtedicilik noktasında iyi çalıştığı kabul edilirken iken, .19 altında değere sahip maddelerin ayırtedicilik noktasında bir daha gözden geçirilmesi önerilmektedir (Ebel & Frisbie, 1986). Öte yandan, point-biserial korelasyonunun 502 katılımcı için hesaplanan sahip olması gereken minimum değer .09'dur (Crocker ve Algina, 1986). Crocker ve Algina (1986) kriterine göre Test 1 içindeki bütün maddelerin ayırtedicilik indeksinin düşük olmasına rağmen, bu değerler hala beklenen (.09 - .3) point-biserial aralığındadır. Her iki ayırt edicilik kriteri dikkate alındığında bir sonraki uygulama için özellikle düşük güçlüğe sahip AP4, LAP14, AP17, VS27, VS18 ve AP10 maddelerinin bir kez daha gözden geçirilmesi gerekmektedir.

Tablo 4.3 Madde istatistikleri (Test 2) (n=506)

	Maddeİstatistikleri			Çeldiriciİstatistikleri (%)					
Madde	*P*	*D*	*r*	A	B	C	D	E	Omit
VP1	0.15	0.14	0.25	0.30	0.32	0.08	0.13	0.15+	0.01
LS6	0.24	0.27	0.45	0.30	0.21	0.13	0.04	0.24+	0.09
AS13	0.19	0.14	0.26	0.42	0.02	0.15	0.19+	0.21	0.01
VS9	0.17	0.18	0.38	0.20	0.17+	0.12	0.18	0.09	0.25
LAP14	0.15	0.19	0.41	0.16	0.25	0.19	0.15+	0.25	0.01
LS11	0.12	0.11	0.26	0.35	0.06	0.29	0.12+	0.15	0.03
AP21	0.23	0.16	0.30	0.08	0.44	0.23+	0.03	0.20	0.02
MS20	0.40	0.36	0.43	0.40+	0.09	0.07	0.17	0.27	0.01
AVP26	0.28	0.18	0.32	0.25	0.28+	0.27	0.09	0.08	0.04
VS27	0.16	0.11	0.20	0.39	0.25	0.16+	0.08	0.11	0.01
LAP7	0.35	0.31	0.41	0.35+	0.15	0.25	0.18	0.06	0.02
LP29	0.32	0.33	0.45	0.14	0.11	0.27	0.16	0.32+	0.02
MS8	0.29	0.22	0.33	0.06	0.40	0.14	0.29+	0.09	0.01
LAS25	0.77	0.27	0.44	0.08	0.03	0.04	0.77+	0.07	0.02
LS16	0.22	0.24	0.37	0.15	0.14	0.22+	0.31	0.07	0.11
AP4	0.29	0.30	0.40	0.13	0.30	0.29+	0.05	0.21	0.03

+: Anahtar

Test 1'e benzer şekilde Test 2 içindeki maddelerin güçlük değerleri .12 ile .77 arasında değerlere sahip olduğu ve Test içindeki 14 maddenin güçlük indeksleri .12 ile .35 arasında değerler aldığı ve genel olarak Test 2 içindeki maddelerin güçlük ortalamalarının ise .27 olduğu gözlemlenmiştir. Test 1'e benzer şekilde test 2 de katılımcılar için zor bir test olarak değerlendirilmiştir.

Test 2 içindeki maddelerin ayırt edicilik indekslerinin sonuçlarına bakıldığında, 506 katılımcı için hesaplanan minimum sahip olması gereken değer .088'dir (Crocker ve Algina,1986) ve bu kritere göre Test 2 içindeki tüm maddeler düşük ayırt edicilik indeksine sahip olmasına rağmen beklenen (.088 - .3) point-biserial aralığındadır. Maddelerden D indeksleri .40 civarında olan üç madde (LAP7, LP29, AP4) bir sonraki uygulamada aynen kullanılmasına, üç maddenin (MS8, LAS25, LS16) çok küçük değişikliklerle tekrarlanmasına, beş maddenin (VP1, AS13, AP21, AVP26, VS27) tekrar gözden geçirilmesine ve kalan son 5 maddenin ise bir sonraki uygulama için elenmesine karar verilmiştir.

Uygulama II sırasında elde edilen verilerin madde analizi sonuçlarına göre Test 1 içindeki toplam 13 madde (LS2, AP3, AP15, AP4, AS19, AP5, AVP26, MS8, VS27, VS18, AP10, LS12, VS24), Test 2 içindeki toplam 11 madde (VP1, AS13, AP21, AVP26, VS27, LAP7, LP29, MS8, LAS25, LS16, AP4) bir araya getirilerek bir sonraki aşamada kullanılmak üzere 20 maddelik Test 3 formu oluşturulmuştur.

4.3. Uygulama III

Üçüncü uygulamanın temel amacı Uygulama II sırasında elde edilen verile ışığında yeni oluşturulan Test formundaki maddelerin Klasik Test Teorisi ve Madde Tepki Kuramı çerçevesinden nasıl çalıştığını ortaya çıkarmaktı. Bu sebeple 20 maddelik yeni set (Test 3) Ankara içindeki iki ayrı üniversiteden 99 kişiden veri toplandı. Toplanan verilerle hem Klasik Test Teorisi hem de RaschAnalizi'ndeki araçlarla madde analizi yapıldı.

4.3.1. Test 3 için Tek Boyutluluk Analiz Sonuçları

ITEMS STATISTICS: MISFIT ORDER

ENTRY	RAW				INFIT		OUTFIT		PTBIS	
NUMBR	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	ITEMS
8	31	97	.03	.22	1.09	1.0	1.20	1.4	A-.01	AVP26
4	23	96	.44	.24	1.13	1.0	1.19	1.0	B-.08	AP4
14	34	95	-.16	.21	1.08	1.0	1.09	.8	C .00	LAP7
11	25	96	.33	.23	1.09	.7	1.03	.2	D-.01	VS24
19	34	80	-.38	.23	1.06	.7	1.08	.7	E .05	LP29
2	73	97	-1.96	.24	1.02	.2	1.07	.4	F .07	AP3
7	42	97	-.49	.21	1.05	.7	1.04	.4	G .05	MS8
16	15	95	.97	.28	1.00	.0	1.02	.1	H .09	VS27
9	41	96	-.45	.21	1.00	.0	1.02	.2	I .14	VP1
17	22	94	.48	.24	1.02	.1	.94	-.3	J .11	LS16
1	42	97	-.49	.21	1.01	.2	1.01	.1	j .12	LS2
6	27	97	.23	.23	.99	.0	.92	-.5	i .13	AP5
13	13	95	1.16	.29	.95	-.2	.99	.0	h .14	AP13
20	6	59	1.69	.42	.98	.0	.76	-.6	g .14	VS18
12	19	96	.70	.25	.98	-.1	.88	-.6	f .16	AP21
10	44	96	-.59	.21	.96	-.5	.96	-.5	e .18	LS12
18	10	88	1.40	.33	.92	-.3	.80	-.6	d .22	AP10
15	59	94	-1.33	.21	.90	-1.3	.90	-.9	c .29	LAS25
5	21	97	.58	.24	.89	-.8	.75	-1.3	b .29	AS19
3	75	95	-2.16	.25	.84	-1.1	.67	-1.8	a .38	AP15
MEAN	33.	93.	.00	.25	1.00	.1	.97	-.1		
S.D.	19.	9.	1.00	.05	.07	.7	.14	.8		

Figür 4.7 Maddelerin tek boyutluluk göstergesi (Test 3)

Rasch Analizi kullanılarak Test 3 içindeki maddelerin tek bir yapısal boyutta olup olmadıkları test edildi. Analiz sonuçları Test 3 içindeki 20 maddenin fit değerlerinin beklenen (0.7 - 1.3) (Linacre, 2007) aralığında olduğu gözlemlenmiştir. Maddelerin ilgili bilgilerinin özeti Figür 4.7'da sunulmuştur.

4.3.2. Test 3 için Madde-Kişi Grafiği

Test 3 içindeki maddelerin kişilerle nasıl eşleştiği, madde güçlük dağılımı, ve kişilerin yetenek kestirimlerine ait bilgiler Figür 4.8'da özetlenmiştir.

```
    PERSONS            MAP OF ITEMS
           <frequ>|<less>
 2                +Q
                  |
                  |
                  |  VS18
                  |
                  |
                  |  AP10
                  |
                  |  AP13
                  |
 1                +S VS27
                  |
               .  |
                 Q|  AP21
                  |  AS19
              .#  |  LS16
                  |  AP4
            .###  |  VS24
                  |  AP5
                  |
 0           ###  +M AVP26
                 S|
            .###  |  LAP7
              .#  |
                  |  LP29  VP1
            ####  |  LS2   MS8
               .  |  LS2
         #######  |
               . M|
               .  |
-1      .#######  +S
                  |
               #  |
           .####  |  LAS25
               .  |
               .  |
               # S|
             .##  |
               .  |
               .  |
-2            .#  +Q AP3
              .#  |
                  |  AP15
                 Q|
               .  |
                  |
                  |
                  |
                  |
-3                +
                  |
                  |
                  |
               .  |
                  |
                  |
                  |
                  |
-4             .  +
            <rare>|<more>
            EACH '#' IS 2.
```

Figür 4.8 Soru ve kişi dağılım göstergesi (Test 3)

Figür 4.8'de görüldüğü üzere maddeler (-3, +3) aralığında dengeli bir dağılım göstermiştir. Bir noktda kümelenme olmaması her bir maddenin farklı bir zorluk derecesine sahip olduğunu göstermektedir. Aynı figür üzerinde en kolay maddenin AP15, en zor maddenin ise VS18 olduğu gözlenebilir. Öte yandan, kişilerin test sonuçlarının dağılımı ise (-4, +1) aralığında olup beklenen (-3,+3) aralığından sapma göstermiştir. Bu sonuçlara göre VS27, AP10, AP13, VS18 maddeleri katılımcıların yeteneklerinin üzerinde kalmıştır. Maddelerin belli denge içinde lineer dağılım göstermesi, sorulan güçlüklerinin belli bir denge içinde dağıldığını göstermektedir. Bu durum ise kişilerin yeteneklerine göre maddelerin ayırt edicilikleri açısından olumlu sonuçlar doğurmaktadır.

4.3.3. Test 3 için Güvenirlik ve Ayırt Edicilik

Figür 4.9 üzerinde özetlenen bilgilerle yola çıkarak Rasch Analizi sonucunda Test 3'e ait madde güvenirlik indeksinin .94 (≥.90) olduğu, maddelerin zorluk olarak yaklaşık 3 kategoriye (with a Separation of 3.80) ayrıldığı görülebilir.

```
                         SUMMARY OF     20 MEASURED ITEMS
+-------------------------------------------------------------------------------+
|           RAW                                MODEL         INFIT       OUTFIT    |
|          SCORE      COUNT      MEASURE    ERROR       MNSQ    ZSTD   MNSQ   ZSTD |
|-------------------------------------------------------------------------------|
| MEAN       32.8       52.3          .00       .25       1.00      .1     .97    -.1 |
| S.D.       18.8        5.0         1.00       .05        .07      .7     .14     .8 |
| MAX.       75.0       55.0         1.69       .42       1.13     1.0   1.20    1.4 |
| MIN.        6.0       34.0        -2.16       .21        .84    -1.3    .67   -1.8 |
|-------------------------------------------------------------------------------|
| REAL RMSE     .25  ADJ.SD      .97  SEPARATION  3.80   ITEM  RELIABILITY  .94 |
|MODEL RMSE     .25  ADJ.SD      .97  SEPARATION  3.85   ITEM  RELIABILITY  .94 |
| S.E. OF  ITEM  MEAN     .23                                                    |
+-------------------------------------------------------------------------------+
```

Figür 4.9 Madde Bilgileri Özeti (Test 3)

4.3.4. Klasik Test Teorisine göre Test 3 Madde İstatistikleri

Klasik Test Teorisinin araçları kullanılarak yapılan analiz sonucunda Uygulama III'ten elde edilen bulgular Tablo 4.4'te özetlenmiştir.

Madde güçlüğü

ITEMAN sonuçlarındaki (p) değerlerini dikkate aldığımızda maddelerin güçlük olarak (p= .05) değerinden (p=.76) değerine kadar farklı değerlerde olduğu gözlemlenmiştir. Güçlük değeri .40'tan daha düşük maddeler zor maddeler olarak tanımlanmıştır (Haladyna, 2004) ve bu bilgiye göre testin geneli katılımcılar için zor maddeler olduğu görülmüştür. İdeal güçlük değeri olan .50 civarında üç madde vardır: AP3 (p= .55), MS20 (p= .44), and LS12 (p= .51). Madde güçlüklerinin ortalaması .30 olması Test 1'in genel olarak zor bir test olduğunu göstermektedir.

Tablo 4.4 Madde istatistikleri (Test 3) (n=99)

Item Statistics				Alternative Statistics (%)					
Item	*p*	*D*	*r*	A	B	C	D	E	Omit
LS2	.42	.43	.51	.42+	.19	.12	.18	.06	.04
AP3	.73	.31	.28	.22	.01	.00	.73+	.02	.02
AP15	.73	.52	.46	.00	.00	.05	.06	.73+	.16
AP4	.20	.23	.23	.38	.03	.20+	.09	.22	.09
AS19	.28	.34	.38	.08	.28+	.50	.03	.07	.05
AP5	.32	.29	.25	.32+	.01	.07	.28	.30	.03
MS8	.48	.34	.37	.06	.24	.18	.48+	.01	.04
AVP26	.25	.29	.25	.28	.36	.04	.25+	.05	.03
VP1	.37	.56	.48	.09	.13	.05	.34	.37+	.03
LS12	.55	.29	.23	.08	.15	.06	.11	.55+	.05
VS24	.22	.30	.32	.16	.11	.22+	.25	.10	.17
AP21	.17	.30	.26	.15	.46	.17+	.02	.18	.03
AS13	.21	.23	.24	.52	.04	.21+	.14	.06	.03
LAP7	.46	.42	.52	.46+	.03	.04	.41	.03	.04
LAS25	.75	.45	.43	.11	.02	.02	.75+	.05	.05
VS27	.15	.37	.42	.18	.18	.24	.15+	.19	.07
LS16	.20	.47	.41	.11	.16	.20+	.31	.07	.16
AP10	.14	.25	.43	.44	.14+	.17	.16	.04	.06
LP29	.43	.32	.27	.26	.02	.10	.15	.43+	.05
VS18	.11	.15	.15	.19	.12	.38	.11+	.13	.08

+: Anahtar

Uygulama III sırasında elde edilen verilerle yapılan analizde ayırt edicilik indeksleri .40 üzerinde olan LS2, AP15, VP1, LAP7, LAS25, ve LS16 maddeleri hem güçlük hem de ayırt edicilik olarak çok iyi çalışmaktadır. Ayırt edicilik indeksleri .30 civarında olan AP4, AS13 ve AP10 maddeler ayırt edici olarak çalışır durumdadır. Madde güçlüğü açısından ise maddeleri (p= .11) ve (p=. 73) aralığında normal bir dağılım göstermiştir (Skewness = .784, andKurtosis = -.486). Test 3'ün güçlük ortalaması.36'dır. Test 1 ve Test 2 ile karşılaştırıldığında Test 3'ün güçlüğü biraz azalmış olsa da Halaydna (2004) kriterlerine göre hala zor bir testtir.

Yukarıda madde indeksleri özetlenmiş olan Test 3 içinde hala çok zor olan VS24, VS18; ayırt ediciliği düşük olan LAP7, doğrudan alan bilgisine yönelik olan LS12 maddeleri bir sonraki uygulama için elenmiştir.

4.4. Uygulama IV

Dördünce ve son uygulamanın temel amacı kalan son 15 maddenin Klasik Test Teorisi ve Madde Tepki Kuramı çerçevesinden nasıl çalıştığını ortaya çıkarmaktır. Bu sebeple 15 maddelik yeni set (Test 4) Ankara içindeki üç ayrı üniversiteden 168öğretmen adayından veri toplandı. Toplanan verilerle daha önceki uygulamalarda olduğu gibi hem Klasik Test Teorisi hem de RaschAnalizi'ndeki araçlarla madde analizi yapıldı.

4.4.1. Test 4 için Tek Boyutluluk Analiz Sonuçları

ITEMS STATISTICS: MEASURE ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	SCORE CORR.	ITEMS
5	10	167	2.39	.31	1.00	.0	1.26	.6	.15	AS19
13	24	165	1.43	.21	1.04	.2	1.10	.4	.19	AP10
7	38	165	.86	.18	1.11	1.0	1.19	1.1	.17	AVP26
11	40	162	.77	.18	.97	-.3	.96	-.3	.32	VS27
12	38	138	.62	.19	1.01	.1	1.01	.1	.28	LS16
6	53	167	.39	.16	.98	-.3	1.01	.1	.34	AS13
8	53	162	.36	.17	.88	-1.7	.80	-1.8	.45	AP21
14	66	163	-.03	.16	1.04	.7	1.04	.5	.30	LP29
4	69	165	-.05	.16	.99	-.3	.96	-.5	.36	AP4
1	69	166	-.06	.16	.87	-2.3	.83	-2.1	.48	LS2
9	70	167	-.07	.16	1.09	1.5	1.11	1.2	.26	VP1
15	88	166	-.55	.16	.96	-.7	.97	-.4	.40	MS8
3	102	167	-.91	.16	1.08	1.2	1.16	1.8	.28	AP4
10	145	166	-2.49	.23	.99	.0	.82	-.8	.39	LAS25
2	126	140	-2.67	.28	.95	-.3	.97	-.1	.39	AP15
MEAN	66.	162.	.00	.19	1.00	-.1	1.01	.0		
S.D.	36.	9.	1.27	.05	.06	1.0	.13	1.0		

Figür 4.10 Maddelerin tek boyutluluk göstergesi (Test 4)

Rasch Analizi kullanılarak Test 4 içindeki maddelerin tek bir yapısal boyutta olup olmadıkları test edildi. Analiz sonuçları Test 4 içindeki 15 maddenin fit değerlerinin beklenen 0.7 - 1.3 (Linacre, 2007) aralığında olduğu gözlemlendi. Maddelerin ilgili bilgilerini özeti Figür 4.10'da sunulmuştur.

4.4.2. Test 4 için Madde-Kişi Grafiği

Test 4 içindeki maddelerin kişilerle nasıl eşleştiği, madde güçlük dağılımı, ve kişilerin yetenek kestirimlerine ait bilgiler Figür 4.11'da özetlenmiştir.

```
     PERSONS MAP OF  ITEMS
            <more>|<rare>
 3                +
                  |
                  |
                  |
                  |T
                  |  AS19
                  |
                  |
 2                +
                  |
                  |
               .  |
                  |
               .  |  AP10
                 T|S
                  |
 1            .#  +
               .  |  AVP26
               .  |  VS27
           #####  |  LS16
               .  |
               # S|  AS13   AP21
          ######  |
                  |
 0             .  +M LS2   LP29  AP5
         #######  |  VP1
               .  |
              ##  |
      .######### M|  MS8
                  |
              ##  |
        ########  |  AP4
-1                +
               .  |
             ###  |S
               . S|
                  |
               .  |
            ####  |
               .  |
-2             .  +
                  |
                 T|
                  |
               #  |T LAS25
                  |  AP15
               .  |
               .  |
-3                +
                  |
                  |
                  |
               .  |
                  |
                  |
                  |
-4             .  +
            <less>|<frequ>
```

Figür 4.11 Soru ve kişi dağılım göstergesi (Test 4)

Figür 4.11’de görüldüğü üzere maddeler (-3, +3) aralığında dengeli bir dağılım göstermiştir. Dha önceki uygulama sonuçlarına benzer şekilde bir noktada kümelenme olmaması her bir maddenin farklı bir zorluk derecesine sahip olduğunu göstermektedir. Aynı figür üzerinde en kolay maddenin AP15, en zor maddenin ise AS19 olduğu gözlenmektedir.. Öte yandan, kişilerin test sonuçlarının dağılımı ise (-

4, +2) aralığında olup beklenen (-3,+3) aralığından sapma göstermiştir. Test 4 için sadece AS19 maddesi katılımcıların yeteneklerinin üzerindedir. Maddelerin belli denge içinde lineer dağılım göstermesi, sorulan güçlüklerinin belli bir denge içinde dağıldığını göstermektedir. Bu durum ise kişilerin yeteneklerine göre maddelerin ayırt edicilikleri açısından olumlu bir sonuçtur.

4.4.3. Test 4 için Güvenirlik ve Ayırt Edicilik

Figür 4.12 üzerinde özetlenen bilgilerle yola çıkarak Rasch Analizi sonucunda Test 4'e ait madde güvenirlik indeksinin .98 (≥.90) olduğu, maddelerin zorluk olarak yaklaşık 6 kategoriye (with a Separation of 6.36) ayrıldığı görülmektedir..

```
     SUMMARY OF     15 MEASURED  ITEMS
+-----------------------------------------------------------------------------+
|          RAW                              MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      66.1     161.7         .00      .19      1.00     -.1   1.01      .0 |
| S.D.      35.7       9.1        1.27      .05       .06     1.0    .13     1.0 |
| MAX.     145.0     167.0        2.39      .31      1.11     1.5   1.26     1.8 |
| MIN.      10.0     138.0       -2.67      .16       .87    -2.3    .80    -2.1 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .20  ADJ.SD    1.26  SEPARATION  6.36   ITEM  RELIABILITY  .98 |
|MODEL RMSE    .20  ADJ.SD    1.26  SEPARATION  6.42   ITEM  RELIABILITY  .98 |
| S.E. OF  ITEM MEAN = .34                                                     |
+-----------------------------------------------------------------------------+
UMEAN=.000 USCALE=1.000
```

Figür 4.12 Madde Bilgileri Özeti (Test 4)

4.4.4. Klasik Test Teorisine göre Test 3 Madde İstatistikleri

Klasik Test Teorisinin araçları kullanılarak yapılan analiz sonucunda Uygulama IV'ten elde edilen bulgular Tablo 4.5'te özetlenmiştir.

Madde güçlüğü

ITEMAN sonuçlarındaki (p) değerlerini dikkate aldığımızda maddelerin güçlük olarak (p= .05) değerinden (p=.76) değerine kadar farklı değerlerde olduğu gözlemlenmiştir. Güçlük değeri .40'tan daha düşük maddeler zor maddeler olarak tanımlanmıştır (Haladyna, 2004) ve bu bilgiye göre testin geneli katılımcılar için zor maddeler olmuştur. İdeal güçlük değeri olan .50 civarında üç madde vardır: AP3 (p= .55), MS20 (p= .44), and LS12 (p= .51). Madde güçlüklerinin ortalaması .30 olması Test 1'in genel olarak zor bir test olarak algılandığını göstermektedir.

Tablo 4.5 Madde istatistikleri (Test 4) (n=168)

Madde İstatistikleri				Çeldirici İstatistikleri (%)					
Madde	*p*	*D*	*r*	A	B	C	D	E	Omit
LS2	.42	.57	.50	.42+	.22	.11	.14	.08	.02
AP15	.84	.45	.55	.00	.00	.01	.02	.84+	.14
AP4	.62	.35	.28	.06	.06	.62+	.14	.12	.01
AP5	.42	.48	.45	.42+	.00	.02	.03	.50	.02
AS19	.28	.45	.44	.06	.28+	.66	.00	.00	.01
AS13	.33	.30	.29	.48	.02	.33+	.14	.02	.01
AVP26	.20	.21	.19	.18	.53	.02	.20+	.04	.02
AP21	.35	.51	.47	.23	.01	.35+	.26	.12	.03
VP1	.42	.26	.28	.09	.06	.28	.42+	.14	.01
LAS25	.96	.13	.43	.01	.00	.01	.96+	.01	.02
VS27	.26	.28	.32	.15	.17	.35	.26+	.03	.04
LS16	.23	.30	.31	.14	.12	.23+	.23	.06	.21
AP10	.14	.16	.21	.39	.14+	.22	.19	.03	.02
LP29	.42	.20	.24	.12	.01	.18	.24	.42+	.03
MS8	.56	.43	.35	.06	.12	.21	.56+	.04	.01

+: Anahtar

Uygulama VI sırasında elde edilen verilerle yapılan analizde ayırt edicilik indeksleri .40 üzerinde olan LS2, AP15, AP5, AS19, AP21ve MS8 ayırt edicilik indeksleri dikkate alındığında bu aşamada çok iyi çalışmaktadır. Ayırt edicilik indeksleri .30üzerinde civarında olan AP4, AS13, ve LS16 maddeler ayırt edici olarak iyi çalışır durumdadır. Ayırt edicilik indeksleri .30 civarında olan VP1 ve VS27 ise çalışır durumdadır. Öte yandan, AVP26, LAS25, AP10 ve LP29maddeleri ayırt edicilik indeksleri açısında çok zayıf kalmışlardır. Testin en kolaymaddesinin LAS25, en zor maddesinin ise AP10 olması bu maddelerin ayırtedicilik indekslerinin düşük çıkmasındaki payı büyüktür.

Test 4 içindeki maddeler (p= .14) ve (p=. 96) aralığında normal bir dağılım göstermiştir (Skewness = 1.15, and Kurtosis = .926). Test 4'ün güçlük ortalaması .43'tür.Halaydna (2004) kriterine göre hala normal bir zorluk değerlerine daha yakındır.

4.5. Geçerlilik Göstergeleri

Test Sonuçları ve Rasch Yetenek Kestirimleri

Katılımcıların Test 4'ten elde ettikleri sonuçlar ile ve kişilerin Rasch yetenek kestirimleri arasındaki ilişki Pearsonproduct-moment korelasyonu yardımı ile incelenmiştir. Analiz öncesinde korelasyon analizinin normallik, doğrusallık ve eş varyanslılıkvarsayımlarında herhangi bir problem gözlemlenmemiştir. Yapılan korelasyon analizi sonucunda kişilerin test sonuçları ile kişilerin yetenek kestirimleri arasında güçlü pozitif bir korelasyon [r=.946, n=144, p=.00] olduğu gözlemlenmiştir.

GPA andRawScores

Katılımcıların Test 4'ten elde ettikleri sonuçları ile üniversitedeki not ortalamaları arasındaki ilişki Pearsonproduct-moment korelasyonu yardımı ile incelenmiştir. Analiz öncesinde korelasyon analizinin normallik, doğrusallık ve eş varyanslılıkvarsayımlarında herhangi bir problem gözlemlenmemiştir. Yapılan korelasyon analizi sonucunda kişilerin test sonuçları ile not ortalamaları arasında ortalama pozitif bir korelasyon [r=.38, n=141, p=.00] olduğu gözlemlenmiştir.

GPA andRaschMeasures

Katılımcıların Rasch yetenek kestirimleri ile üniversitedeki not ortalamaları arasındaki ilişki Pearsonproduct-moment korelasyonu yardımı ile incelenmiştir. Analiz öncesinde korelasyon analizinin normallik, doğrusallık ve eş varyanslılık varsayımlarında herhangi bir problem gözlemlenmemiştir. Yapılan korelasyon analizi sonucunda kişilerin yetenek kestirimleri ile not ortalamaları arasında ortalama pozitif bir korelasyon [r=.40, n=141, p=.00] olduğu gözlemlenmiştir.

BÖLÜM 5

TARTIŞMA VE SONUÇ

Bu çalışma kapsamında ilköğretim matematik öğretmen adaylarının ölçme kavramlarını özellikle uzunluk, alan ve hacim ölçme kavramlarını öğretme bilgilerini ölçmeye yönelik çoktan seçmeli bir test geliştirilmiştir. Geliştirilen testin analiz sonuçları aday öğretmenlerin matematik öğretme bilgilerini bir aşamaya kadar ölçebildiğini göstermiştir.Shulman (1986, 1987) çalışmalarında tamamen teorik olarak açıkladığı pedagojik alan bilgisinin uygulamada bir karşılığının olduğunun sinyallerini almak, bundan sonraki test geliştirme çalışmaları adına motive edicidir.

Bu çalışmanın ana amacı matematik öğretmen adaylarının ölçme kavramlarını öğretme bilgilerini yönelik çoktan seçmeli bir test geliştirmekti. Bu ana amaç doğrultusunda öğretmen adaylarının özellikle hangi kavram yanılgılarına sahip oldukları, nelerde zorluk yaşadıklarını çok detaylı şekilde ortaya çıkarmak için başka bir araştırma desenine ihtiyaç olduğundan bu çalışma kapsamında elde edilen bulgular gözlem niteliğinde kalmış, sonuçlar bölümünün ilgili yerlerinde kısa notlar halinde raporlanmıştır. Bu gözlemler ve bu gözlemlere dayanarak matematik eğitimi açısından sunulan öneriler şu şekilde sıralanabilir.

Çok temel olarak matematik öğretmen adaylarının ölçme konusundaki kavramsal bilgilerle ilgili ciddi sıkıntılar yaşadıkları gözlemlenmiştir. Ölçme kavramlarının altında yatan matematiksel düşünceleri anlamakta, diğer matematiksel kavramlarla ilişkilendirmekte, matematiksel formülleri anlamlandırmakta ciddi güçlükler yaşamaktadırlar. Bu durumun sadece ölçme ile sınırlı olmadığı veöğretmen adaylarının ve hatta çalışan öğretmenlerin öğrencilerle aynı zorluk ve kavram yanılgılarına sahip oldukları yine literatürdeki bilgiler arasındadır(örn. Even, 1990;vanDriel, 1998, Ward, 2004). Literatürdeki bilgilere benzer şekilde, bazı öğretmen adaylarının test maddeleri içinde kurgulanan öğrenci zorluklarını birebir yaşadıkları ve kendilerinden bir öğretmen olarak müdahale etmesi beklenen durumlarda zayıf kaldıkları gözlemlenmiştir.

Öğretmen adaylarının yaşadıkları temel matematiksel bilgi eksikliklerinin yanı sıra,birebir görüşmelerde öğretmen adayları, öğrencilerin düşünme yaklaşımları ile ilgili deneyime sahip olmadıklarını ve almış oldukları dersler kapsamında bu bilgileriniçok fazla sorgulamadıklarını ifade etmişlerdir.Buna bağlı olarak öğretmenlik bilgilerini kullanmaları gereken üst düzeydeki test maddelerinde daha fazla zorluk çektikleri çalışma kapsamındaki gözlemlerden bir diğeridir. Geliştirilen testle ilgili yapılan analizlerde doğrudan matematiksel bilgi içeren maddelerin zorluk ve ayırt ediciliklerinin daha iyi çalıştığı, pedagojik alan bilgisine yönelik hazırlanan maddelerin ise güçlük ve ayırt edicilik indekslerinin beklenen değerlerden daha düşük olduğu gözlemlenmiştir.

Oysaki Shulman (1986) çalışmasında pedagojik alan bilgisinde iki ana vurgu yapmıştır. İlki, pedagojik alan bilgisinin "bir kavramı başkaları için anlaşılabilir yapmak için en işlevsel fikir, analoji, gösterim formül gibi temsil şekilleri" (s.9) içerdiğini söylemektedir. Bu noktada kişinin bir başkasına bir fikri işlevsel bir formda aktarabilmesi için öncelikle kişinin kendisinin o kavramla ilgili fikirlerinin net ve kapsamlı olması gerekmektedir. İkincisi ise, pedagojik alan bilgisine sahip bir kişinin öğrencilerle ilgili herhangi bir kavramı öğretirken o kavram öğretimini kolaylaştıran veya zorlaştıran etmenler ve öğrencilerin hali hazırda sahip oldukları kavramlar ve ön kavramlar ile ilgili bilgilere sahip olması gerektiğini söylemektedir (Shulman, 1986, s.9).

Bu amaçla,öğretmen adaylarının öğrenciliklerinden beri getirdikleri zorluk ve kavram yanılgılarının lisans eğitimleri sırasında giderilmesi gerekmektedir. Öğretmen adaylarına matematiksel kavramlar üzerinde düşünmeleri için fırsatlar tanınmalı, özellikle temel matematiksel kavramlar ve bu kavramların öğretimine yönelik derslere önem verilmelidir. Matematiksel kavramların anlamlarının tartışılmasının yanı sıra kavramlar arası ilişkilendirme, çoklu gösterimler ve matematiksel düşüncelerin doğrulama ve ispat çalışmalarına yer verilmelidir.

İkinci olarak öğretmen adaylarının öğrencilerle ilgili bilgilerinin artırılması gerekmektedir. Öğretmen adayları öğrencilerin karakteristiklerinin yanı sıra karşılarına çıkan bir öğrenci durumunda öğrenciyi anlayabilmeli, öğrencinin yaşadığı zorluğu ya da kavram yanılgısını tespit edebilmeli, bu durumun kaynağı hakkında fikir sahibi olabilmelilerdir.Eğitim fakültelerinden mezun olmadan önce öğretmen adaylarına literatürlerde tespit edilmiş öğrenci karakteristikleri, yatkınlıkları, konu bağlamında yaşadıkları zorluklar ve kavram yanılgıları verilmelidir. Bir diğer öneri ise, yine öğretmenlik bilgisini güçlendirmeye yönelik olarak farklı öğretim metotlarına yer verilmeli, öğretmen adaylarının duruma daha geniş bir acıyla bakabilmesi sağlanmalıdır.

Bir diğer öneri ise öğretmen eğitimi boyutundadır. Yapılan ara/alt analizler sırasında farklı üniversitelerdeki öğretmen adaylarının farklı cevaplama örüntüleri oluşturduğu gözlemlenmiştir. Fakat, öğretmen adaylarının neden bu şekilde cevaplama örüntüleri oluşturduğunu ortaya çıkarmak bu çalışmanın amacı dışında olduğu için detaylı olarak raporlanmamıştır. Üniversiteler arasında bu şekilde farklılıkların ortaya çıkmasında özellikle özel öğretim derslerinin farklı içeriklerde sunuluyor olması ve öğretmen adaylarının öğretmenlik bilgilerinin oluşumu sırasında bir standardizasyon olmadığından kaynaklanıyor olabilir. Veri toplama çalışmaları sırasında özel öğretim yöntemleri derslerini veren öğretim üyeleri ile görüşmeler yapılmış, yapılan görüşmeler sırasında tasarlanan ders içerikleri, kullanılan kaynak ve yöntemlerin farklılık gösterdiği gözlemlenmiştir. Bu amaçla yüksek öğretim kurumunun belirlemiş olduğu akreditasyon çalışmalarının üzerine gidilmesine, her eğitim fakültesinde öğretmen adaylarına benzer deneyimlerin sunulması gerekmektedir. Ayrıca öğretmen adaylarının bu şekilde cevap örüntüsü oluşturmasının altında yatan nedenleri bilimsel olarak ortaya çıkarmak için farklı desenlerde uzun soluklu çalışmaların yapılmasına ihtiyaç vardır.Eğitim fakültelerinde bu çalışma kapsamında geliştirilen test ve geliştirilebilecek benzer testlerinbelli aralıklarla uygulanmasını içerenuzun soluklu çalışmaların yapılması mümkündür. Belli aralıklarla uygulanan testlerin sonuçlarıeğitim fakültelerdeki öğretim programları ve öğretmen eğitimi için çok önemli bilgiler sunacak ve öğretmenlik bilgisinin oluşma sürecindeki kritik zaman dilimlerinin ortaya çıkarılması sağlayacaktır.

Bundan sonraki öğretmenlik bilgisine yönelik test geliştirme çalışmaları için söylenebilecek öneriler şu şekilde özetlenebilir. Araştırmacılar pedagojik alan bilgisinin yapısı hakkında henüz ortak bir görüş oluşturmuş değillerdir. Hatta yapılan çalışmalarda yapısına dair kabul gören bazı doğrulanamadığı ortaya çıkmıştır. Örneğin, tek boyutlu olduğu varsayılan öğrenciler ve alan bilgisinin tek boyutlu olmadığı Schilling (2007) çalışmasında raporlanmıştır. Bu belirsizlik durumutest geliştirirken soru yazımından, analizlerin belirlenmesine kadar pekçok alanda araştırmacıların zorluk yaşamasına neden olsa da Pedagojik alan bilgisinin kristalize olması ve netleşmesi için bu tarz çalışmalara ihtiyaç vardır. Test geliştirme çalışmalarının sonuçları öğretmenlik bilgisine ait teoriyi, teorideki değişiklikler ölçme araçlarının evrilmesini sağlayacaktır.

Kulikowich ve Alexander(1994) öğretmen ve öğrencilerin testlerde verdikleri yanıtların rasgele olmadığını, bu cevaplama yaklaşımın altında belli başlı kavram yanılgısı, zorluk gibi kimi bilgiler barındırabileceğini ifade etmiştir. Doğru-yanlış şeklinde puanlanan test maddelerinin bu bilgileri tespit etme noktasında zayıf kalacağı aşikardır. Her ne kadar ikili puanlama türünün diğer puanlama türlerine göre psikometrik özelliklerolarak daha avantajlı olduğu bilinse de, özellikle öğretmenlik bilgisi gibi oldukça kapsamlı bir alana hitap eden testler için özel yapılandırılmış çeldiricilerle dereceli puanlama yaklaşımın bundan sonraki çalışmalarda daha kapsamlı bilgiler sunacağı düşünülmektedir.

Son olarak, bu çalışma öğretmenlik bilgisine yönelik test geliştirme çalışmalarının öncülerinden olan bu çalışmada elden geldiğince test maddelerinin işlerliğini görmek adına elden geldiğince kapsamlı bir çalışma yürütülmüştür. Test içindeki kimi maddelerin psikometrik değerleri, beklenen değerlerin altında kalmış olsa da, bu alanda yapılacak çalışmalar için umut vadedici sonuçlar elde edilmiştir.