

# Effects of Content-Associated Applied Measurement and Evaluation Course

## Alanla İlişkilendirilmiş Uygulamalı Ölçme ve Değerlendirme Dersinin Etkileri

Gülden Gürsoy<sup>1</sup> , Mustafa Aydoğdu<sup>2</sup> 

<sup>1</sup>Faculty of Education, Adıyaman University, Adıyaman, Turkey

<sup>2</sup>Faculty of Education, Gazi University, Ankara, Turkey

### Özet

Bu araştırmanın amacı, alan ile ilişkilendirilmiş uygulamalı Ölçme ve Değerlendirme dersinin öğretmen adaylarının ölçme ve değerlendirme okuryazarlıklarına, ölçme ve değerlendirme dersine yönelik tutumlarına ve alan bilgilerine etkisini incelemektir. Araştırma ayrıca, deney grubu öğretmen adaylarının, alanla ilişkilendirilmiş ölçme ve değerlendirme uygulamalarına yönelik görüşlerini ortaya koymayı hedeflemektedir. Araştırmada karma yöntem olarak adlandırılan, nicel ve nitel araştırma yöntemlerinin birlikte kullanıldığı araştırma modeli kullanılmıştır. Araştırmanın nicel boyutu, 90 fen bilgisi öğretmen adayı ile nitel boyutu ise, deney grubunda yer alan 6 öğretmen adayı ile yürütülmüştür. Araştırmada veri toplama aracı olarak; Ölçme ve Değerlendirme Okuryazarlığı Ölçeği, Ölçme ve Değerlendirme Dersine Yönelik Tutum Ölçeği, Akademik Başarı Testi ve Odak Görüşme Formu kullanılmıştır. Araştırmanın sonucunda, alanla ilişkilendirilmiş ölçme ve değerlendirme uygulamalarının öğretmen adaylarının ölçme ve değerlendirme okuryazarlık düzeylerini geliştirdiği, ölçme ve değerlendirme dersine yönelik tutumlarını olumlu yönde etkilediği ve konu alan bilgi düzeylerini artırdığı belirlenmiştir. Öğretmen adayları ile yapılan odak görüşme, hem nicel sonuçların nedenini açıklamış hem de öğretim sürecinde uygulanan yöntemin hizmet etme derecesine ilişkin ayrıntılar vermiştir.

**Anahtar sözcükler:** Konu alan bilgisi, öğretmen yetiştirme, ölçme ve değerlendirme dersine yönelik tutum, ölçme ve değerlendirme okuryazarlığı.

### Abstract

The objective of the present study is to examine the effects of content-associated applied measurement and evaluation course on pre-service teachers' assessment literacy, their attitudes towards the measurement and evaluation course, and their field-specific content knowledge. Furthermore, the study aims to reveal the pre-service teachers' views on content-associated applied measurement practices. The mixed method research approach, which combines qualitative and quantitative research methods, was used in the study. The quantitative data collection was conducted with 90 science teacher candidates. The qualitative data collection was conducted with six teacher candidates from the experimental group. The Classroom Assessment Literacy Inventory (CALI), Measurement and Evaluation Course Attitude Scale (MECAS), Academic Achievement Test (AAT), and Focus Group Interview Form were used as the data collection tools in the study. It was found that the content-associated applied measurement and evaluation course improves the measurement and evaluation literacy of the candidate teachers, fosters their positive attitudes towards the course of measurement and evaluation, and increases their content knowledge. Furthermore, the focus group interviews with the pre-service teachers both helped explain the reasons behind the quantitative results, and provided detailed information about the degree of effectiveness of the method applied in the teaching process.

**Keywords:** Assessment literacy, attitude towards the course of measurement and evaluation, content knowledge, teacher education.

In the world education system, the emergence of standards-based educational movements and the widespread adoption of accountability policies in education has improved the measurement and evaluation capacities that are expected of teachers in the measurement and evaluation applications conducted in the classroom during recent years (Deluca & Klinger, 2010; Volante & Fazio, 2007). The measurement and evaluation applications that were conducted to classify the successful or unsuccessful students were replaced by the contemporary

measurement and evaluation approaches that play an active role in the instruction process, demonstrate the degree of learning, and sustain learning in the process (Stiggins, 2008). Teachers are expected to possess the competencies to develop or select measurement instruments, manage the evaluation approaches, provide feedback, grade, and share the results accurately with stakeholders (McMillan, 2014; Popham, 2011; Russell & Airasian, 2012), as well as the ability to assess whether the students acquired the lifelong learning skills such as critical think-

#### İletişim / Correspondence:

Gülden Gürsoy  
Faculty of Education, Adıyaman  
University, Adıyaman, Turkey  
e-mail: gakdag@adiyaman.edu.tr

Yükseköğretim Dergisi / Journal of Higher Education (Turkey), Çevrimiçi Erken Baskı / Online Preprint Issue. © 2019 Deomed  
Geliş tarihi / Received: Mayıs / May 17, 2018; Kabul tarihi / Accepted: Ocak / January 23, 2019

Bu çevrimiçi makalenin atf künyesi / Please cite this online article as: Gürsoy, G., & Aydoğdu, M. (2019). Effects of content-associated applied measurement and evaluation course. *Yükseköğretim Dergisi*, doi:10.2399/yod.19.004

This article is produced from the PhD thesis entitled "Influence of the Course of Practical Assessment and Evaluation Associated with the Field on the Literacy Levels of Assessment and Evaluation Attitudes and Content Knowledge of Teacher Candidates".

ORCID ID: G. Gürsoy 0000-0002-4886-7645; M. Aydoğdu 0000-0002-1504-3674



ing, problem solving, decision making skills, entrepreneurship, sense of responsibility, collaboration and communication skills (Andrade, Du, & Wang, 2008; Brookhart, 2011; Pellegrino & Hilton, 2012). The changes in the measurement and evaluation process have altered the measurement and evaluation competency standards for teachers both in developed countries (USA, UK, Canada) and in Turkey (Volante & Fazio, 2007).

Initially, the measurement and evaluation standards were determined by the American Federation of Teachers (AFT), National Council on Measurement in Education (NCME) and National Educational Association (NEA) in 1990, and these standards were developed by Stiggins (1991) as “Assessment Literacy”. Teachers who meet these standards were referred to as “assessment literate”. Stiggins (1991) defined the assessment literate individual as one who is aware of the difference between objective or subjective assessments. As the qualifications of teachers changed along with the educational reforms, Stiggins defined the assessment literate individual as an individual who knows why the assessment is conducted, who can select / develop a measurement instrument, interpret the measurement data, decide after the process, and report the results. Several studies (Davies, 2008; Fulcher, 2012; Mertler & Campbell, 2005; Popham, 2011; Walters, 2010) have attempted to define assessment literacy and the behavior of an assessment literate teacher. The said studies indicate that assessment literate individuals have a better command of their classes, know the objective of the assessment, could select or develop the optimal measurement instrument, make sure the measurement tools are valid and reliable and assess the measurement results objectively. Furthermore, these studies indicate that assessment literate individuals are aware of the potential harms of an inaccurate assessment and the fact that their calculations or interpretations of the assessment results could influence the decisions on both the student and the education.

However, as findings of several studies (Beziat & Coleman, 2015; Davidheiser, 2013; Gotch & French, 2013; Mertler, 2003; Mertler & Campell, 2005; Tao, 2014; Volante & Fazio, 2007; Yamtim & Wongwanich, 2014) and the studies conducted in Turkey (Birgin, 2007; Gül, 2011; Karaman, 2014; Karaman & Şahin, 2014; Ogan-Bekirođlu & Suzuk, 2014) established, determination of the qualifications required to become an assessment literate teacher could not guarantee that the teacher would be assessment literate. Based on this fact, studies in the literature started to concentrate on the development of assessment literacy. Some studies examined the correlation between literacy levels and demographics (Tao, 2014; Zhang & Burry-Stock, 1997), while others examined the cor-

relation between the literacy and attitudes of in-service and pre-service teachers (Quilter & Gallini, 2000). Some went a step further and studied the development of assessment literacy levels (Karaman, 2014; Mertler, 2009; Tsagari, 2008). In general, the significance of assessment literacy has been recognized worldwide and the topic has been studied extensively.

However, this global trend has not been reflected in the studies conducted in Turkey and there are only a few studies on the methods to develop the assessment literacy of teachers in Turkey (Buldur, 2009; Karaman, 2014; Ogan-Bekirođlu & Suzuk, 2014; Özsevgeç, Çepni, & Demirciođlu, 2004; Uluman, 2009). In the present study, a method was developed based on the potential factors (measurement and evaluation course, attitude, topical content knowledge) that would affect the problems experienced in the achievement of assessment literacy by pre-service teachers. The main research question was whether the said method that was implemented in the measurement and evaluation course improved the assessment literacy of pre-service teachers.

The problems in the content or instruction of the measurement and evaluation course, where the initial basic concepts, thoughts and attitudes are formed, are the basis of the inadequate self-perception of teachers or their low literacy levels in measurement and evaluation. The most significant problem in the failure of the comprehension of the topics in the course was the lack of practical applications in the theoretical and comprehensive measurement and evaluation course that prevents the application of the content learned by the students (Volante & Fazio, 2007), and rapid coverage of the topics due to the intense content in the course curriculum. The students perceive the theoretical and rapid instruction as complex, developing a negative attitude towards the measurement and evaluation course or its application. Given the fact that attitudes shape behavior, the problem worsens (Brown, 2008; Inbar-Lourie & Donitsa-Schmidt, 2009; Quilter & Gallini, 2000).

The present study aimed to design a learning environment where pre-service teachers could learn how to apply their theoretical knowledge into practice to improve their assessment literacy levels. Furthermore, the learning environment included activities which demonstrated that the measurement and evaluation course was not complicated and was fun, and proved that learning could take place while performing these activities. The activities included in the present study were expected not only to improve the cognitive aspects of pre-service teachers' assessment literacy levels, but also develop their affective and psychomotor skills.



## The Objective of the Study

The objective of the present study was to examine the effects of a content-associated applied measurement and evaluation course on pre-service teachers' assessment literacy and their attitudes towards measurement and evaluation course, and on their content knowledge. Furthermore, the study aimed to reveal the views of pre-service teachers on the process.

The following questions were asked in the research.

- Is there a significant difference between the pre-experimental and post-experimental assessment literacy and attitudes towards the measurement and evaluation course between experimental group pre-service teachers who were instructed with content-associated applications and control group pre-service teachers who were instructed with traditional methods?
- Is there a significant difference between the pre-experimental and post-experimental learning domain achievements of the experimental group pre-service teachers who were instructed with content-associated activities?
- What are the views of experimental group pre-service teachers on content-associated measurement and evaluation applications?

## Methodology

### Research Method

In this study, the mixed method was used in which both quantitative and qualitative methods were used together.

Mixed method studies are not simple combinations of qualitative and quantitative methods, but comprehensive and integrated studies that utilize the strengths of both methodologies to support each other (Creswell, 2014; Firat, Kabakçı-Yurdakul & Ersoy, 2014; Greene, 2007; Johnson & Onwuegbuzie, 2004; Teddlie & Tashakkori, 2009). The explanatory design was selected from among various mixed methods.

An explanatory design consists of first collecting quantitative data and then collecting qualitative data to help explain or elaborate on the quantitative results. The rationale for this approach is that the quantitative data and results provide a general picture of the research problem; but more analysis, specifically through qualitative data collection is needed to refine, extend or explain the general picture (Creswell, 2014).

The quantitative data were collected with pretest-posttest control group quasi-experimental design, a semi-experimental model, and pre-test posttest weak experimental model. The qualitative data were collected with focus interview model to investigate the emotions, thoughts and perceptions that could not be determined through quantitative data collection instru-

ments after the experimental process and concurrent with the posttest applications based on the explanatory mixed design. The findings were interpreted by comparisons.

Since there were two classes attending the “Measurement and Evaluation” course at the university where the research was conducted and the grade average of the students in both classes were similar, one of the existing classes was designated as the experimental and the other was designated as the control group in the study. In the study, the measurement and evaluation course was instructed with content-associated measurement and evaluation applications to the experimental group, while the control group was instructed with the conventional educational approach without such an association. Furthermore, a field achievement test was applied as the pre-test and post-test to investigate whether practicing in the field during the application activities increased the experiment group's content knowledge. The experimental designs used in the study are presented in ■ Table 1.

■ Table 1. The experimental design.

	Pretests	Experimental process	Posttests
Experimental group	CALI MECAS AAT	Content-associated applied measurement and evaluation activities	CALI MECAS AAT
Control group	CALI MECAS	Measurement and evaluation activities conducted with conventional instructional approach	CALI MECAS

AAT: Academic Achievement Test on Matter and Change/States of Matter Units; CALI: Classroom Assessment Literacy Inventory; MECAS: Measurement and Evaluation Course Attitude Scale.

## Participants

In the present study, since generalization was not required, a purposive sampling method, criterion sampling was used to ensure data diversity. Since the focus of the study was the content of the measurement and evaluation course, the science pre-service teachers who attended this course were selected on purpose. Adiyaman University Faculty of Education was chosen as the research site for its easy accessibility.

Two types of participants were selected to be included in the study since the study aimed to collect both quantitative and qualitative data. In collection of the quantitative study data, 90 pre-service science teachers ( $N=90$ ), who attended the “Measurement and Evaluation” course in formal education and evening education groups at Adiyaman University Faculty of Education, participated in the study. The grade point averages of the participating students for the previous semester were

obtained from the student automation system and the comparability between the classes was checked before the application of the pretest. It was determined that the grade point average of the pre-service science teachers attending formal education was 2.73 and the grade point average of the pre-service science teachers attending evening education was 2.87. Due to the similarity between the grades of the groups, one group (evening education) was randomly selected as the experimental group ( $N=45$ ; 33 females and 12 males) and the other group (formal education) was assigned as the control group ( $N=45$ ; 29 females and 16 males).

The second part of the study included the qualitative dimension. In this part, only volunteering students participated in the study. A classification was conducted for the maximum variation sampling method, which is particularly used in studies where the data is collected with interviews. In maximum variation sampling, the aim is to maximize the diversity of the individuals who participate in the study, and to reflect this fact to the study as much as possible. The main objective is to determine whether there are common or shared phenomena among diverse situations, rather than making generalization possible (Yıldırım & Şimşek, 2008). Thus, six volunteering pre-service teachers were selected in the experimental group and a focus interview was conducted with these pre-service teachers. In order to assure data diversity, the pre-service teachers with good, moderate and poor academic achievement levels were selected. The pre-service teachers Fuat (very good-male), Kader (good-female), Sevda (moderate-female), Burcin (moderate-female), Ceylan (poor-female) and Peyami (very poor-male) were selected for the focus interviews.

### Data Collection Instruments

The data collection instruments are given below.

#### Classroom Assessment Literacy Scale (CALI)

It was decided that the scale developed by Mertler and Campbell (2005) would be adequate to use in the study since it is both current and developed particularly for pre-service teachers. Mertler and Campbell (2005) developed a 35-item scale, which included five scenarios and seven questions in each scenario, consistent with the standards of measurement and evaluation competence established by AFT, NEA and NCME. It was determined that the mean item difficulty was .64, mean item discrimination was .32, and measurement reliability was .75. Bütüner, Yiđit and Çimer (2010) translated the scale developed by Mertler and Campbell (2005) to Turkish language and conducted the validity-reliability study. The mean item difficulty of the scale was .64, mean item discrimination was .48, and the measurement reliability was .86.

Bütüner et al. (2010) stated that the scale is capable of measuring the measurement and evaluation knowledge and skill levels of the teachers and pre-service teachers based on the measurement validity and reliability.

The CALI was applied once before the experimental procedure and once more after the experimental procedure. 90 students including the students in the experimental and control groups participated in the application. The measurement reliability of the scale in the present study was .74. In order to make the pre-service teachers better relate to the scenarios, the courses in the scenario were replaced with the Science and Technology course.

#### Measurement and Evaluation Course Attitude Scale

“Measurement and Evaluation Course Attitude Scale,” developed by Aktaş and Alicı (2012), was used in the study. The scale included three factors, 15 positive and five negative items. The first factor included eight items, the second factor included seven items and the third factor included five items. The scale is a five-point Likert type tool and scored as “I completely disagree,” “I disagree,” “Neutral,” “I agree,” “I completely agree.” The measurement reliability of the scale was determined as .944. The measurement reliability of the scale in the present study was .738.

#### Achievement Test on Matter and Change Learning Domain

The present study aims to investigate the effect of the measurement and evaluation course by association to the field of science on the content knowledge levels of pre-service science teachers in the experimental group. Thus, an achievement test was developed to determine the content knowledge. The learning domain that would be included in the achievement test was randomly selected by the author from the “Matter and Change” learning domain in the 8th grade Science Curriculum, which was updated in 2013. An achievement test that included multiple choice questions, which measured the knowledge level based on 8th grade achievement indicators was developed. The steps identified in the literature for the development of an achievement test were strictly implemented.

When developing the achievement test on “Matter and Change”, the achievement indicators were entered in the column section and the target skill for each outcome was noted by the researcher and the advisor individually. In order to ensure the content validity of the achievement test, it was decided to include at least 24 test items in the achievement test based on the achievement indicators. The table was finalized after the review by two curriculum development experts.



In the study, commercially available TEOG exam question repositories were reviewed to select the questions adequate for the achievement indicators after the label table was finalized. It was determined that 23 test items were needed in order to ensure content validity of the achievement test; however, it was decided that two questions that test each achievement had to be included in the achievement test since the item analysis revealed that the items did not serve the purpose adequately. Thus, 49 test items related to the achievement indicators were developed using the question repositories. The preliminary application of the achievement test was conducted with 104 freshmen and sophomore students attending the General Chemistry Course at the Adiyaman University Faculty of Education Science Teaching Department. The exam papers of 104 students were evaluated with the SPSS 16 software.

The item analysis that assisted the item selection was conducted after the item-score matrix was determined. Item analysis included two phases. The first was the item discrimination index that also ensures the validity of the test. The second was the item difficulty index.

Fifteen items with an item discrimination index of .28 in the preliminary application of the test were excluded (items 1, 3, 6, 8, 11, 13, 18, 22, 27, 32, 37, 39, 44, 47, 49). Of the remaining 34 items, the item with the highest discrimination index was selected from the two test items that tested the same achievement indicators, and the achievement test that included 26 items was finalized. When there was more than one item with an adequate discrimination index value for a behavior, the item with the highest discrimination index value can be selected (Kan, 2008). Most test items should be moderately difficult. The reason for the selection of items with moderate difficulty is the fact that such items possess the maximum item variances. Maximum item variance means that the item can better reflect the differences between individuals for the trait that needs to be measured (Kan, 2008). The mean difficulty of the items in the 26-item achievement test was .49, and the mean discrimination index was .44. After this process, test analyses were conducted based on the final test scores and the analysis results are presented in Table 2.

Cronbach's alpha coefficient was calculated. The test measurement reliability (Cronbach's coefficient) was determined as

.719. According to Kalaycı (2008) and Özdamar (1999), the scale has high reliability. Thus, it could be argued that the measurements obtained in the achievement test were reliable.

### Focus Interview

Focus interview technique was selected since it fit well with the study objective. The questions were designed based on the categories of introduction, overview, transition, key, research, epilogue, and final questions. Attention was paid to make sure that the questions were comprehensible, short, clear and with a single objective. The questions were ordered from the general to particular questions. In order to test the comprehensibility of the questions, they were read aloud to the two pre-service teachers in the experimental group and the problematic points were revised. Two faculty members at Adiyaman University, who previously conducted qualitative research, were asked whether the questions were appropriate. Thus, after the necessary editing was conducted, the audio recorder and the interview venue were determined. A setting with a round table, where each participant's voice could be heard was arranged.

Six volunteering pre-service teachers were selected from among the pre-service teachers in the experimental group for the focus interview. In order to determine the appropriate time, the teacher candidates were interviewed one by one and the members of the group convened on the appropriate day and hour. The content was explained to the students and it was stated that the data will be used without mentioning their names. The interview lasted approximately one hour. The moderator was the researcher and tried to give each participant enough time to speak, and to focus the discussion on the topics when the discussion digressed from the topic at hand.

The focus interview questions are given below:

- What are the things you liked about the measurement and evaluation course?
- What are the things you did not like when practicing the things you learned in the measurement and evaluation course?
- Do you think that the practical applications conducted in the measurement and evaluation course contributed to your learning?
- Did developing an achievement test change your content knowledge, and do you think teacher's level of content knowledge is important when developing the questions?

### Data Collection

The study was conducted in the Measurement and Evaluation Course, a sixth semester course in the Science Teaching Program. Prior to the study, the Measurement and Evaluation

Table 2. Achievement test analysis results.

N	Mean	SD	Median	Mode	Mean diff.	Mean discrimination	Cronbach's alpha
104	12,32	4,41	13	15	.49	.44	.719



Course instructor was informed about the application content and permission for instruction by the researcher was requested. After obtaining the permission, a written request was sent to the department head and administrative permission was also obtained.

Since the pre-service teachers started their classes on the second week of the semester after the holidays, instructions started on the second week as well. Furthermore, the ALS and measurement and evaluation course attitude scale were administered as the pretest on the same week.

For seven weeks, the measurement and evaluation course was delivered by the researcher based on the content in the academic information set to both the experimental and the control groups. After the seventh week, the experimental process was initiated for about six weeks. For six weeks, statistics included in the undergraduate measurement and evaluation course curriculum, namely, item and test statistics, test development, assessment and grading, statistical procedures on the measurement results, and development of grade scoring keys were taught to the students in the control group. As homework, the students were asked to solve several problems. In the experimental group, these topics were taught with applied methods. Initially, the researcher administered the academic achievement test as a pre-test before starting the treatment with the experimental group.

### Experimental Process

During the first week of the experimental process, the pre-service teachers were asked to design an achievement test related to “Matter and Change” learning domain similar to the academic achievement test designed by the researcher. Pre-service teachers were reminded about the steps to design such a test and these steps were discussed with the pre-service teachers. When a consensus was reached, the researcher distributed the papers and the tables that contained the achievement indicators to pre-service teachers. Pre-service teachers were required to prepare the table of specifications in the same class. The achievement indicators were entered in the table of specifications based on Bloom’s taxonomy. In the second class, whether these indicators were distributed into the accurate steps was discussed with the class and checked. After deciding how many questions to prepare in total, the researcher distributed the question bank that included 15 questions to the pre-service teachers and asked them to Xerox-copy the questions in certain sections to utilize these questions. They were asked to construct the questions within a week.

In the second week of the experimental process, it was observed that certain pre-service teachers were not able to design the achievement test. During the course of that class, the achievement tests of those who designed the achievement test were reviewed individually and they were asked to correct their mistakes if any. One more day was allowed to complete the missing achievement indicators and for students who had not designed the test. Four days later, a meeting was held again for approximately one hour and all the pre-service teachers submitted the designed achievement tests they had designed to be checked by the researcher.

In the third week of the experimental procedure, the achievement tests that had been reviewed by the researcher for scope and face validity were brought to the class and the problems in the submitted tests were shared with the pre-service teachers. The most elaborate achievement test was introduced to the class and explained why it was chosen. A copy of the achievement test was distributed to the pre-service teachers and they were allowed to create the answer keys within a class-hour. For one class-hour, the responses of the pre-service teachers and the actual answers were discussed, and the achievement tests were finalized. The pre-service teachers were then allowed to form groups of 3-5 individuals and to give the test to 8th grade students in any school. The researcher arranged the schools where the tests would be given to certain groups.

In the fourth week of the experimental process, the students who brought back the administered tests were asked to score the tests using the answer key prepared the previous week. Meanwhile, the researcher frequently reminded students about the rules that should be followed when scoring the tests. In order to ensure the scoring reliability of the tests pre-service teachers were allowed 20 minutes to discuss the scores with their group members. After a consensus was reached, a random test was selected, and the results were copied for the statistical process.

In the fifth week of the experimental procedure, the item analysis phase was conducted. In order to construct the item-score matrix, 1 point was assigned for each correct response of the student for each item, and 0 point was assigned for incorrect or blank responses. The item-score matrix of the most successful 27% (upper group) and the least successful 27% (lower group) were analyzed based on the internal criterion. Each pre-service teacher calculated the discrimination and difficulty indices for each item in the test. After the calculations were over, the items that would be excluded were decided in the classroom based on the criteria in the literature. To finalize the tests, whether a question that corresponds to each achievement



indicator existed was controlled and the item variance, standard deviation and reliability coefficients were calculated.

In the sixth week of the experimental procedure, test statistics were conducted on the findings. The pre-service teachers ranked the test scores in an ascending order and prepared the frequency tables. The of central tendency measures including the mode, median, mean, etc. of the measurement findings and the measures of central dispersion including range, standard deviation, and variance were calculated. The dispersion curve was plotted after the measurement process was completed. Finally, the Z and T scores were calculated for each question. It was not possible to complete the course within allotted time, thus, two days later, the students were called to the class and the poster design, vocabulary association and graded scoring keys applications were conducted. Furthermore, a chat on the student projects and homework assignments was held regarding the “Matter and Change” topic. AAT was applied to experimental group students as the posttest. Only during the finals week, the posttest was given to all pre-service teachers.

During the first month of the senior year spring semester, focus interviews were conducted with the pre-service teachers selected for the focus interview on the previously determined date and time. Thus, the experimental process was finalized. Then, the analysis of the collected data was conducted.

### Data Analysis

Shapiro-Wilks (SW) and skewness-kurtosis coefficients were used to determine whether the data demonstrated a normal distribution in the study. Shapiro-Wilks is utilized when the group size is smaller than 50, while Kolmogorov-Smirnov (KS) is used when it is greater than 50 (Büyüköztürk, Çakmak, Akgün, Karadeniz, & Demirel, 2009). Shapiro-Wilks test and skewness-kurtosis coefficients demonstrated that the experiment and control group scores in the “CALI, MECAS and AAT”, exhibit normal distribution. A higher than .05 *p*-value in the Shapiro-Wilks test corresponds to normal distribution of the data. The distribution is considered normal when the skewness and kurtosis coefficients are between -1.96 and +1.96. Furthermore, the Levene test was performed to determine the homogeneity of the variances. When the *p*-value obtained in the Levene test is higher than .05, then the group variances are considered to be homogeneous. The Levene test was used to control the homogeneity of the group variances, and Box’s M statistic was used for covariance homogeneity. Two-factor ANOVA test for mixed measures was used to determine whether the mean test and scale scores of pre-service teachers in different groups dif-

fered before and after the application due to the utilized instruction methods. Non-parametric Mann-Whitney U test was used for the MECAS.

The use of descriptive analysis technique, one of the qualitative data analysis methods, was considered to be more suitable for the analysis of data obtained in the focus interviews. Descriptive analysis is one of the qualitative analysis techniques that include direct citations to reflect the views of the interviewed or observed individuals (Yıldırım & Şimşek, 2008).

The analysis demonstrated that pre-test and post-test scores in ALS and AAT displayed normal distribution. However, the pretest scores obtained from the measurement and evaluation course attitude scale did not exhibit normal distribution. It was decided that nonparametric tests should be used for data without normal distribution and parametric tests should be used for the data with normal distribution.

In the study, the focus interviews were recorded with an audio recorder. Immediately after the interviews, the recordings were transcribed. After about five hours of labor, a 35-page document was obtained. The answers of each pre-service teacher were read several times and their common views were interpreted by the researcher. Furthermore, the most significant view that represented the common opinion was directly quoted. In the last stage, the findings were compared with the results found in the literature.

### Validity and Reliability

In the study, the time interval between the pretest and the posttest was determined as six weeks in order to ensure internal validity and prevent the participants from remembering their pretest answers. Furthermore, to improve the internal validity of the study, the focus interview questions were reviewed by two experts. After the assessment by these experts, two questions were excluded, and one question was edited. In order to ensure the external validity of the study, the experimental methodology and the focus interview stages are explained in detail. Furthermore, maximum diversity sampling technique based on purposive sampling was used to select the participants. In order to ensure the validity of the research, attention was also paid to ensure that the data collection period was in ideal length required by the focus group study. In order to improve the reliability of the study, the measurement reliability of the scales used in the study was calculated based on the posttest scores. Furthermore, assistance of an independent researcher, who was experienced in the focus interview method, was obtained for the interpretation of the data collected via the focus interviews.

## Findings

### Quantitative Findings

Two-way ANOVA for mixed measures (ANOVA) was used in the analysis of the experimental and control group pretest and posttest data to determine the effect of content-associated measurement and evaluation course on the assessment literacy levels.

The descriptive data on the experimental and control group pretest and posttest scores are presented in ■ Table 3.

The experimental and control group CALI pretest and posttest mean scores and standard deviations are presented in ■ Table 3. Table data demonstrated that the pretest mean score of the pre-service teachers in the CALI ( $M=10.82$ ) was very close to that of the pre-service teachers in the control group ( $M=9.38$ ). The posttest mean score of the pre-service teachers in the experimental group ( $M=17.06$ ) was higher than the posttest mean score of the pre-service teachers in the control group ( $M=12.06$ ). The mixed measures two-way ANOVA test was used to determine whether the difference between the pretest and the posttest scores was significant. The mixed measures two-way ANOVA test are presented in ■ Table 4.

- There was a significant difference between the experimental and control group pretest and posttest total scores before and after the experimental process [ $F_{(1-88)}=60.216; p<.05$ ]. This finding suggests that the assessment literacy levels of pre-service teachers in the experimental and control groups differed regardless of the measurement applied.
- There was a significant difference between the experimental and control group assessment literacy mean pretest and posttest scores [ $F_{(1-88)}=133.012; p<.05$ ].
- It was found that the common effect of the factors that reflected being in different application groups and measurements conducted at different times on the assessment literacy skills of the pre-service teachers was significant [ $F_{(1-88)}= 21.071; p<.05$ ]. In other words, the method used in the experimental group was more effective in improving the assessment literacy scores of the pre-service teachers.

Cohen's  $d$  formula was used to determine whether there was a significant difference between the posttest scores of the

■ Table 4. ANOVA test findings on achievement pretest and posttest scores.

Source of the variance	KT	SD	KO	F	p
Between subjects	1150.000	89			
Group	467.222	1	467.222	60.216	.000
Error	682.778	88	7.759		
Within subjects	1634.000	90			
Measure (Pre/Post)	897.800	1	897.800	133.012	.000
Group* measure	142.222	1	142.222	21.071	.000
Error	593.978	88	6.750		
Total	2784.000	179			

\*Cohen's  $d=1.72; \eta^2=.430$ .

pre-service teachers in the experimental group, where the measurement and evaluation course was taught through the content-associated approach, and the control group, where the instruction was conducted through the conventional instruction approach. As a result, it was found that the  $d$ -value was 1.72. Thus, it was determined that the content-associated approach had a strong effect on the assessment literacy levels of pre-service teachers (Cohen, 1988).

In order to determine the effect of the content-associated measurement and evaluation course on the attitudes of the teacher candidates towards the measurement and evaluation course, the analysis of the pretest and posttest attitude scores of the experimental and control groups was conducted via the Mann-Whitney test, a nonparametric test, since the data did not exhibit normal distribution. The two-way ANOVA for mixed measures was not used, since it requires normal distribution and would not provide reliable results, and instead the Mann-Whitney U test was preferred.

The descriptive data regarding the experimental and control group pretest and posttest scores are presented in ■ Table 5.

Based on the table, the attitude scale pretest mean score of the pre-service teachers in the experimental group ( $M=3.62$ ) was very close to the attitude scale mean score of the pre-service teachers in the control group ( $M=3.64$ ). The Mann-Whitney U test, one of the nonparametric tests, was applied

■ Table 3. Descriptive data on literacy scale pretest and posttest scores.

Tests	Groups	N	M	SD
Pretest	Experimental	45	10.82	2.47
	Control	45	9.38	2.44
Posttest	Experimental	45	17.06	2.79
	Control	45	12.06	3.02

■ Table 5. Descriptive data on attitude pretest and posttest scores.

Tests	Groups	N	M	SD
Attitude pretest	Experimental	45	3.62	.60
	Control	45	3.64	.63
Attitude posttest	Experimental	45	4.09	.438
	Control	45	3.48	.585



■ **Table 6.** Mann-Whitney U test results on the comparison of MECAS pretest scores of the groups.

Tests	Groups	N	M	S.T	U	Z	p
Pretest	Experimental	45	3.61	2039,000	1.004	-.069	.945
	Control	45	3.64	2056,000			

to determine whether the difference between the experimental and control group pretest scores was significant. The test results are presented in ■ Table 6.

Analysis of the ■ Table 6 demonstrated that there was no significant difference between the measurement and evaluation course attitude scale pretest mean scores ( $M=3.61$ ) of the pre-service teachers in the experimental group and the MECAS pretest mean scores of the pre-service teachers in the control group ( $M=3.64$ ) ( $Z=-.069$ ;  $p>.05$ ). The lack of a difference between the attitude scores of the pre-service teachers towards the measurement and evaluation course before the experimental process revealed more information about the effectiveness of the experimental process.

The mean posttest MECAS score of the pre-service teachers in the experimental group ( $M=4.09$ ) was higher than that of the pre-service teachers in the control group ( $M=3.48$ ). The only difference between the mean pretest and posttest MECAS scores of the pre-service teachers was observed in the experimental group. Independent groups t-test, a parametric test, was conducted to determine whether this difference between the posttest scores of the experimental and control groups was significant. The test findings are presented in ■ Table 7.

Analysis of ■ Table 7 demonstrated that there was a significant difference between the mean MECAS posttest score of the pre-service teachers in the experimental group ( $M=4.09$ ) and mean MECAS posttest score of the pre-service teachers in the control group ( $M=3.48$ ) ( $t=5.518$ ,  $p<.05$ ).

Cohen's  $d$  formula was used to calculate the effect size of the abovementioned analysis result, which revealed that the  $d$ -value was 1.17. The effect size obtained for this analysis sug-

■ **Table 7.** Independent groups t-test results on the comparison of MECAS posttest scores of the groups.

Tests	Groups	N	M	Ss	SD	t	p
Posttest	Experimental	45	4.09	.438	.88	5.518	.000
	Control	45	3.48	.585			

$$\text{Cohen } d = 5.518 \sqrt{45+45/45*45} = 1.17$$

gested that the content-associated applied instruction had a high level of effect on changing the attitudes of pre-service teachers towards the measurement and evaluation course (Cohen, 1988).

The dependent groups t-test was conducted to determine whether the affinity of the pre-service teachers with the field during the applications that were taught with content-associated approach in the measurement and evaluation course had an effect on the improvement of their content knowledge. The descriptive data on the pretest and posttest achievement scores for the experimental group are presented in ■ Table 8.

The AAT pretest and posttest scores of the experimental and control groups are presented in ■ Table 8. Based on the table, the mean pre-test AAT score of the pre-service teachers in the experimental group ( $M=13.66$ ) were similar to their posttest academic achievement score ( $M=17.82$ ). In order to determine whether the difference between the pretest and posttest mean scores of the experimental group was significant, dependent groups t-test, a parametric test, was applied. The test results demonstrated that there was a significant difference between the academic achievement pretest scores ( $M=13.66$ ) and posttest scores ( $M=17.82$ ) of the pre-service teachers in the experimental group ( $t=6.241$ ;  $p<.05$ ).

■ **Table 8.** Descriptive data on pretest and posttest AAT scores.

Group	Test	N	M	SD	T	p
Experimental	Pretest	45	13.66	3.52	6.241	.000
	Posttest	45	17.82	2.79		

## Qualitative Findings

### Findings on the First Question

Six students participated in the focus interview that was conducted to determine student views on the measurement and evaluation course. Each student was assigned a pseudonym for anonymity purposes. The male pre-service teachers were coded as Fuat and Peyami, while the female pre-service teachers were coded with the nicknames Ceylan, Kader, Burcin and Sevda. Some direct quotes from the student views expressed in the focus interview were presented and interpreted. The findings related to each question were addressed under separate themes.

### Findings on the First Question

The answers given by the pre-service teachers to the question "What are the things you liked about the measurement and evaluation course?" generally indicated satisfaction. It can be



stated that they were satisfied with the course delivery although at different degrees. For example, while Ceylan was pleased with the topical discussions in the class, Peyami stated that he was more satisfied with the process of designing the exam questions. In particular, they enjoyed the process of cutting and pasting the questions selected according to the specification tables from different sources.

*Ceylan: The moment I enjoyed the most, for instance, in the class, the presence of such a discussion environment. The fact that the course was instructed in the form of a brainstorm. For example, while you asked questions, I was very pleased that we were producing answers.*

*Peyami: The part where we designed the questions was fun. You select the questions after you master the topic. It was fun for me to cut and paste.*

Pre-service teachers designed an achievement test that included different measurement and evaluation techniques on the topic of “matter and change”. Then, an achievement test, which was decided by class consensus, was given to the students at schools either identified by them or by the researcher. After the test administration process, the item analysis was conducted on the questions and the reliability of the test was calculated. The pre-service teachers mostly expressed their satisfaction about the said application. In particular, Fuat, Kader and Ceylan stated that as teachers during the practical applications and that they felt happy to interact with the students.

*Fuat: I enjoyed learning by doing and living. Because it was different from other courses, I actually went and implemented the test in a class myself. The thing I liked the most was actually implementing the test at the school. We were actually interacting with the students.*

*Kader: Similarly, the fact that I actually implemented the test with the students at the school. We provided the students with the questions, we functioned as a teacher, we solved the tests together with the students.*

Sevda and Burcin indicated that they enjoyed it the most when they scored the achievement tests of the students and calculated the item difficulty and discrimination levels for each item.

*Sevda: I enjoyed finding out for myself the validity and reliability of the problem I developed in the process. I liked conducting the item analyses.*

The pre-service teachers liked the cut-and-paste process when developing the achievement test, giving the achievement test to the students and scoring the tests and conducting the item analysis the most, during which the pre-service teachers perceived themselves as teachers. It seems that the students

liked these practical applications since they made the pre-service teachers feel important and improved their self-awareness.

Encountering the facts about their future profession even for a short period of time could result pre-service teachers’ self-awareness and positive development of their attitudes towards their profession.

### Findings on the Second Question

The pre-service teachers were not unanimous on the question “What are the things you did not like when practicing the things you learned in the measurement and evaluation course?” While Fuat stated that he did not experience any difficulties during the practical applications, the other pre-service teachers stated the distinct problems they experienced. For example, Burcin stated that she had difficulties in finding the adequate questions for the cognitive classifications in the specifications table that was developed based on the achievement indicators for the achievement test and was not happy about it.

*Burcin: When I was preparing the questions after the table of specifications, I experienced a little difficulty. I had difficulties in finding questions adequate for the level of the students.*

The problem experienced by the pre-service teacher Burcin might be due to her indecision about the adequate cognitive step for the questions in the test, which was probably induced by lack of sufficient resources, student’s lack of knowledge on how to utilize the internet for that purpose, or on the classification of the questions based on Bloom’s taxonomy.

Ceylan complained about the time it took to complete the practices in the class and the fact that she had to work at home in certain occasions. This could be related to the her negative attitude towards doing the tasks related to school at home. Having an idea that schoolwork should stay at school might have caused this dissatisfaction with the abovementioned situation.

*Ceylan: It took too much time, we had to work at home and at school. But when we realized that this was productive, for example, even now, when we see these courses, we realize it. But we were a little exhausted.*

Peyami stated that during the reminders about the previous class provided at the beginning of the next class, he sometimes felt disconnected with the course and that bothered him. The fact that this process did not continue during practical applications can be considered as one of the limitations of the present study. The fact that the students exploited this was ignored during the study.

*Peyami: And there was also this fact: When you were conducting basic instruction, you asked questions about the topics taught in the previous class and reviewed these topics. For example, in my experi-*



mental group, I was absent during the last few weeks. These topics, I had no idea about. The next time I attended the class, the previous topics were not kind of reviewed. This may be because there were no instructions during the hands-on applications.

Sevda stated that she was not pleased with the requirement of going to a school to conduct the achievement test. This can be related to the fact that the pre-service teacher was uncomfortable with interacting with the social environment or with the actual conditions (spatial distance, health problems, financial difficulties).

Sevda: *There was a lot of things that I really liked when participating in such a practice. The only thing I disliked was having to actually go to a school to implement these and then read them all and conducting calculations. I did not like the idea of going to a school. It was very tiring.*

Kader stated that the anxiety he experienced at the beginning of the hands-on applications caused problems and he was not satisfied with that. This may be due to the fact that the content was not made clear by the researcher in a way that everyone could understand it clearly.

Kader: *At first, you divided us as experimental and control groups, we were actually scared. We thought that we would get tired more in the experiment group and worried about our productivity. But then we realized that it was not so bad when we put it into practice, and even realized that we learned better.*

### Findings on the Third Question

Except for one pre-service teacher, all the pre-service teachers responded to the question “Do you think that hands-on applications that we conducted in the measurement and evaluation course contributed to your learning?” by stating that the application process enabled them to learn the topics in the measurement and evaluation course better, they found out that they remembered the most of what they learned when they reviewed the topics during the PPSE preparation process, the hands-on applications they conducted facilitated the interpretation of the questions and they were able to interpret the definitions or topics they did not know based on what they knew.

Sevda: *The hands-on applications were a great contribution to us. When we were preparing for PPSE, we did not have to study in detail. Because at that stage, I learned almost all the field content. When directly attempted to solve the tests, I could remember it all. It enabled us to learn them permanently. Maybe, if it was theoretical knowledge, it would depend on memorization. I would have forgotten it after a certain period of time, but since we did it that way, I learned more by myself. This provided permanent learning.*

Fuat: *It really was very permanent for me. For example, we can even come up with definitions. We know what measurement, evalu-*

*ation means. For example, we remember the formulas at a glance. Maybe this was the biggest difference between us and our friends in the control group. For example, rather than trying to learn the formulas, we deducted the formulas through calculations, so the learning was more permanent.*

The pre-service teachers Ceylan, Sevda and Kader stated that they did not experience any difficulties in remembering the measurement and assessment course during the PPSE preparation and solved the tests without spending much time reviewing the topics. This can be an indication that the learning occurred during the comprehension and application phase. Permanent learning is one of the expected conditions for a teacher during the instruction of a course. It can be argued that permanent learning in fact occurred based on the statements of the pre-service teachers. It can be stated based on the statements by Fuat and Burcin that learning took place in higher cognitive steps when solving problems. The fact that they could even interpret the questions on the topics they did not know and were able to come up with the definitions they did not know based on the definitions they already knew demonstrated that another objective of the course was achieved.

Peyami stated that the theoretical knowledge taught during the first seven weeks of the measurement and evaluation course was more permanent. In particular, he emphasized that he learned the topics better during the sections where the researcher asked the students to think and discuss the topic in the classroom. Thus, although it contradicts with the principle of learning by doing, this could be due to the different learning style of the particular pre-service teacher or the problems he experienced during the application process.

Peyami: *With me, it was just the opposite. For example, when I was studying for PPSE, the information you taught with the conventional method were more permanent for me. Because you were asking questions in the preliminary section in the class. We were preparing for it in advance because we knew that you would ask questions. Therefore, more permanent learning took place. The practice section was more towards the professional life.*

### Findings on the Fourth Question

It was determined that the pre-service teachers shared similar views about the questions “Did developing achievement test change your content knowledge?” and “Is teacher’s level of content knowledge important when developing the questions?” For example, Ceylan stated that while she developed the achievement test, her knowledge on the topic improved and she realized her shortcomings when solving the questions developed by classmates. In fact, this was desirable in the study process. The present study aimed to create awareness among



students about their topical content knowledge and to follow a developmental process on the issue.

Ceylan: *For me, us attempting to solve the problems that were designed by our classmates in the classroom and to observe our shortcomings were the greatest advantage, for example, when you research for the questions, the things you feel you do not know well, the topics that I was not well, I thought that 'Yes, I lack that information' and I decided to start doing something differently. I mean, the course in general contributed to me.*

Burcin and Sevda stated that they reviewed the “matter and change” unit before designing the achievement test and then they started to develop the questions. This suggested that the pre-service teachers were aware of the need for topical knowledge to design a question and detailed knowledge on the topic was needed to develop higher level questions. Furthermore, Burcin used expressions, which indicated that the learning could take place both while developing and solving problems. This could support the fact that the difference between the “Matter and Change” learning domain achievement test pretest and posttest scores were significant in the experimental group in the quantitative section of the study.

Burcin: *As my friends stated, when I was developing the questions, I realized that I was missing certain information, and there were certain small points that I had to learn. I investigated a topic first, then inquired how to design questions in that topic. Because I did not think it was appropriate to design a low-level question. After I investigated the topic, I got help from other resources and as such designed the test, and I think that I developed my content knowledge. I think this helped me, I have a higher degree of knowledge in sciences now.*

Peyami stated that he had trouble with his level of content knowledge while developing the questions. This enabled the pre-service teacher to recognize his lack of content knowledge and review the topic. Pre-service teachers Kader and Fuat expressed in more detail why they should have content knowledge when designing the questions. The knowledge that easy, intermediate and difficult questions should be asked to differentiate the students with and without knowledge and to make this possible, the teacher should master the subject led pre-service teachers Kader and Fuat to review the subject. Designing questions is a part of the synthesis dimension, which is among the high-level learning skills. The fact that pre-service teachers reviewed the topic before designing the questions could indicate that the instruction of content knowledge courses in teacher training institutions are not adequate for the development of high-level learning skills.

Peyami: *In order to design questions, it is necessary to master the subject. In the 8th grade topics, we were already checking out the pre-*

*viously designed questions when designing our own. We were inspired by these questions. I have noticed that I did not know several topics. I felt my knowledge was not sufficient in that field. This made it difficult for me to develop the questions. First of all, we studied the subject, the achievement indicators and then designed the questions.*

Fuat: *We had to develop questions at every level. We had to develop easy-intermediate and also difficult questions. Perhaps we could have developed easy questions without too much content knowledge. Perhaps we could also develop the intermediate questions, but to differentiate the students, to design the harder questions, we had to absolutely master the subject. So, after observing the achievement indicators, we had to review the topic. Thus, we developed the multiple-choice and other questions.*

## Conclusions and Discussion

The findings of the present study demonstrated that the method implemented in the instruction of the measurement and evaluation course affected the assessment literacy levels of pre-service teachers. Previous study findings are consistent with the present study and demonstrate that there is a close correlation between the instruction method implemented in measurement and evaluation course and the assessment literacy level of the students (Beziat & Coleman, 2015; Karaman, 2014; King, 2010; Mertler, 1999; Mertler, 2009; Tzagari, 2008; Uluman, 2009; Wise, Lukin, & Roos, 1991). Mertler (2009) state that the measurement and evaluation course supported by workshop activities improved teachers' measurement and evaluation skills, and teachers could transfer the knowledge to the students more rapidly thanks to these hands-on applications. Cohen and Hill (2000) state that positive changes were observed in measurement and evaluation practices conducted by teachers in schools where workshops are conducted once or twice per year, and such activities that are conducted only a few days per year could improve assessment literacy of the teachers. Discussion of the achievement indicators for student achievement in department meetings would benefit the teachers attending these meetings because they will receive feedback from other teachers. Furthermore, such applications would contribute to the professionalization of teachers in the field of assessment and evaluation in schools (Shepard et al., 2005). A study conducted by Karaman (2014), determined that the measurement and evaluation activities conducted with micro-instruction approach contributed to the development of assessment literacy of the pre-service teachers by providing in depth knowledge on measurement and evaluation methods, improving their interpretation ability, and enabling feedback and reflection.

Ainsworth and Viegut (2006) found a positive correlation between student achievement and teachers' assessment liter-



acy levels. According to DeLuca and Klinger (2010), the promotion of the theoretical dimension of the measurement and evaluation course with hands-on applications would contribute to the development of the assessment literacy of pre-service teachers.

Development of the measurement and evaluation course content is the responsibility of the instructors of the course and the previous studies indicate that the level of knowledge and skills of the instructors on measurement and evaluation affect the assessment literacy levels of pre-service teachers (Green & Stager, 1986; Stiggins, 2010; Xu & Liu, 2009). The method applied in the present study is expected to assist the future instructors of the measurement and evaluation course.

Giving the pre-service teachers opportunities to design graded scoring keys for the project, poster and homework topics, and to construct answer keys for the achievement test that they designed made it possible for them to observe, for the first time, how the measurement and evaluation process was conducted to measure their own academic achievement. This also motivated them to have a higher level of achievement based on clear assessment criteria and disambiguation. Previous studies (Wayman, Midgley, & Stringfield, 2006; Wiliam, Lee, Harrison, & Black, 2004) demonstrate that even a single measurement-evaluation practice with clear criteria could positively affect student learning and achievement.

One of the results obtained in focus interviews confirmed the improvement in assessment literacy levels. At the end of the applied measurement and evaluation course, the pre-service teachers indicated that the knowledge they acquired was permanent, they remembered the content without a need to review it again and they gained experience on how to implement their knowledge.

In general, previous studies demonstrated that the content of the measurement and evaluation course, where the foundations of the measurement and evaluation competency that every individual who perform the teaching profession should possess are laid, should be supplemented with extra materials and hands-on practice.

It was also found in the present study that the content-associated and applied measurement and evaluation course positively changed the attitudes of the pre-service teachers towards the assessment and evaluation course. Attitude and achievement are two elements that interact with each other. The attitude developed towards a course or program constitutes 25% of the achievements in that course or program (Bloom, 1976). Previous studies (Green & Stager, 1986; Quilter & Gallini, 2000; Richardson, 1995) revealed the presence of a correlation between attitude and achievement. For teachers to achieve the

expected competencies on measurement and evaluation, first they need to have a positive attitude towards measurement and evaluation. It can be argued that positive development of teachers' attitudes towards measurement and evaluation is dependent on the quality of the measurement and evaluation course. In an undergraduate teacher education classroom where the desired level of learning cannot be attained, the pre-service teacher would acquire insufficient professional knowledge and skills and this would prevent the pre-service teacher's predisposition to be trained as a good teacher (Fishbein & Ajzen, 2010). Quilter and Gallini (2000) argue that a teacher with a negative attitude towards measurement and evaluation practices does not include the required measurement and evaluation practices in her or his class. Bonner and Chen (2009) state that teachers with negative attitudes towards measurement and evaluation do not have the ability to interpret measurement-evaluation data. Richardson (1995) states that teachers' attitudes towards measurement and evaluation constantly interact, and professional development can only be achieved through a change in these two elements. Thus, it is necessary to examine the pre-existing ideas of teachers on measurement and evaluation during training to improve their assessment literacy (Brown, 2008).

The findings of the present study demonstrate that the content-associated and applied measurement and evaluation course positively changed the attitudes of the pre-service teachers towards the measurement and evaluation. Other studies also present similar arguments (McMillan, 2001). The inclusion of practical activities in the measurement and evaluation courses that the pre-service teachers attend would enable them to start their profession with a more advanced approach (Şahin & Karaman, 2013). Furthermore, it can be argued that the positive attitudes of pre-service teachers towards the measurement and evaluation course will increase the possibility of performing the activities that they conducted during their undergraduate studies in their own classes in the future (Fishbein & Ajzen, 2010; Quilter & Gallini, 2000).

The findings obtained in focus interviews supported the positive change in attitudes. The cut-and-paste activities conducted while designing the achievement test, the fact that they assumed the role of the teacher and conducted exams and graded the papers made the measurement and evaluation course fun and the students enjoyed themselves. Furthermore, the fact that pre-service teachers realized that the analyses conducted on the test results were not as difficult as they had expected also encouraged them to conduct measurement and evaluation practices.



The present study also demonstrates that the content-associated and applied measurement and evaluation course increased the field content knowledge of the pre-service teachers. Pre-service teachers repeatedly reviewed the topics in the unit during the process. It can be argued that they achieved better learning since they noticed the details in every topic they investigated as a professional teacher. The interviews reveal that the pre-service teachers' content knowledge improved when designing the achievement test in particular. They stated that they could not find the questions that measure the high level skills during the question design process according to the Bloom's taxonomy, and they had to review the topics to be able to design the questions. They indicated that their content knowledge of chemical industry improved significantly. It can be argued that the fact that content knowledge of the pre-service teachers improved as they conducted measurement and evaluation activities also improved their measurement and evaluation skills, a conclusion supported by the previous studies.

Studies that investigated the correlation between pedagogical content knowledge and field content knowledge of teachers demonstrate that content knowledge has a positive effect on pedagogical content knowledge (Alkharusi, Kazem, & Al-Musawai, 2011; Canbazoglu, Demirelli, & Kavak, 2010; Duncan & Noonan, 2007; Halim & Meerah, 2002; Jadama, 2014; Käpylä, Heikkinen, & Asunta, 2009; Türnüklü, 2005). In a study by Halim and Meerah (2002) that investigated the pedagogical content knowledge of 12 teachers on selected physics concepts, it was found that the pedagogical knowledge of the teachers with inadequate content knowledge was also inadequate. Studies reported that pre-service teachers without content knowledge or with inadequate content knowledge could not properly teach students, leading to the emergence of conceptual misconceptions in students (Halim & Meerah, 2002; Hashweh, 1987; Stacey et al., 2001). Alkharusi et al. (2011) determined that the pedagogical courses associated with subject-specific courses maximized the professional skills of the pre-service teachers. Lederman and Gess-Newsome (1992) acknowledge that content knowledge affected in-class applications, however note that these two factors are in fact interrelated.

The focus interview findings support the quantitative findings. It can be stated that the application of theoretical knowledge learned in the measurement and evaluation course positively affected the development of assessment literacy levels of pre-service teachers, the fact that they interacted with the students while conducting the hands-on application at schools positively affected the development of their attitudes, and the design of the achievement test and topics for project

and homework assignments positively affected the development of their content knowledge.

## Recommendations

Below, several recommendations are presented based on the findings of the current study.

The content-associated applied activities conducted in the present study contributed to the development of assessment literacy levels of the pre-service teachers. Based on the study results, it can be suggested that faculty members could implement this method in their classes.

In the present study, the measurement and evaluation course was divided into equal periods, where the theoretical and practical applications were conducted together. It may be advisable to conduct measurement and evaluation courses with a method where the two are delivered in more or less balanced time periods, rather than only including the theoretical or practical approach.

A second measurement-evaluation course, where pre-service teachers can practice the theoretical knowledge they acquired, could be included in the curriculum and instructed in the semester that follows the initial course such as "Measurement and Evaluation II".

There is no activity related to measurement and evaluation in the 14-week activity program for School Practice and Teaching courses. Such activities may be conducted by the counselors at schools if the timeframe is insufficient for the faculty members to conduct the hands-on applications included in the present study. It is advisable to include some hands-on measurement and evaluation activities in the activity program for the School Practice and Teaching course.

## References

- Ainsworth, L., & Viegut, D. (2006). *Common formative assessments: How to connect standards-based instruction and assessment*. Thousand Oaks, CA: Corwin.
- Aktaş, M., & Alici, D. (2012). Eğitimde ölçme ve değerlendirme dersine yönelik Tutum Ölçeği'nin (EÖD-TÖ) geliştirilmesi. *Qafqaz Üniversitesi Dergisi*, 33, 63-73.
- Alkharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal of Teacher Education*, 39(2), 113-123.
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association (1990). *The standards for teacher competence in the educational assessment of students*. Accessed through <<http://files.eric.ed.gov/fulltext/ED323186.pdf>> on January 13, 2014.
- Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice* 27(2), 3-13.



- Beziat, T. L. R., & Coleman, B. K. (2015). Classroom assessment literacy: Evaluating pre-service teachers. *The Researcher*, 27(1), 25–30.
- Birgin, O. (2007). Sınıf öğretmenleri adaylarının ölçme-değerlendirme konusundaki okuryazarlık düzeylerinin incelenmesi. *XVI. Ulusal Eğitim Bilimleri Kongresi*, 5–7 Eylül 2007, Tokat.
- Bloom, B. S. (1976). *Human characteristics and school learning*. New York, NY: McGraw-Hill.
- Bonner, S. M., & Chen, P. P. (2009). Teacher candidates' perceptions about grading and constructivist teaching. *Educational Assessment*, 14(2), 57–77.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3–12.
- Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York, NY: Nova Science Publishers.
- Buldur, S. (2009). *Fen bilgisi öğretmen adaylarının alternatif ölçme ve değerlendirme yaklaşımlarına yönelik okuryazarlık ve öz yeterlik düzeylerinin geliştirilmesi*. Yüksek lisans tezi, Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü, Sivas.
- Bütüner, S. Ö., Yiğit, N. ve Çimer, S. O. (2010). Ölçme değerlendirme okuryazarlığı envanterinin Türkçe'ye uyarlanması. *E-Journal of New World Sciences Academy*, 5(3), 792–809.
- Büyükoztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2009). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi.
- Canbazoglu, S., Demirelli, H., & Kavak, N. (2010). Fen bilgisi öğretmen adaylarının maddenin tanecikli yapısı ünitesine ait konu alan bilgileri ile pedagojik alan bilgileri arasındaki ilişkinin incelenmesi. *İlköğretim Online*, 9(1), 275–291.
- Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record*, 102(2), 294–343.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches*. Thousand Oaks, CA: Sage.
- Davidheiser, A. S. (2013). *Identifying areas for high school teacher development: A study of assessment literacy in the Central Bucks School district*. Doctoral thesis, Drexel University, School of Education, Philadelphia, PA, USA.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3) 327–347.
- DeLuca, C., & Klinger, D. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Assessment in Education*, 17(4), 419–438.
- Duncan, C. R., & Noonan, B. (2007). Factors affecting teachers' grading and assessment. *The Alberta Journal of Educational Research*, 53(1), 1–21.
- Fırat, M., Kabakçı-Yurdakul, I., & Ersoy, A. (2014). Bir eğitim teknolojisi araştırmasına dayalı karma yöntem araştırması deneyimi. *Eğitimde Nitel Araştırmalar Dergisi*, 2(1), 65–86.
- Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York, NY: Psychology Press (Taylor & Francis).
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132.
- Gotch, C. M., & French, B. F. (2013). Elementary teachers' knowledge and self-efficacy for measurement concepts. *The Teacher Educator*, 48(1), 46–57.
- Green, K. E., & Stager, S. F. (1986). Measuring attitudes of teachers toward testing. *Measurement and Evaluation in Counseling and development*, 19, 141–150.
- Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, LA: Jossey-Bass.
- Gül, E. (2011). *İlköğretim öğretmen adaylarının ölçme-değerlendirme okuryazarlığı ve ölçme-değerlendirmeye ilişkin tutumlarının belirlenmesi*. Yüksek lisans tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ.
- Halim, L., & Meerah, S. M. (2002). Science trainee teachers' pedagogical content knowledge and its influence on physics teaching. *Research in Science & Technological Education*, 20(2), 215–225.
- Hashweh, M. Z. (1987). Effects of subject matter knowledge in the teaching of biology and physics. *Teaching & Teacher Education*, 3(2), 109–120.
- Inbar Lourie, O., & Donitsa Schmidt, S. (2009). Exploring classroom assessment practices: The case of teachers of English as a foreign language. *Assessment in Education: Principles, Policy & Practice*, 16(2), 185–204.
- Jadama, L. M. (2014). Impact of subject matter knowledge of a teacher in teaching and learning process. *Middle Eastern & African Journal of Educational Research*, 7, 20–32.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher* 33(7), 14–26.
- Kalaycı, Ş. (Ed.). (2008). *SPSS uygulamalı çok değişkenli istatistik teknikleri*. Ankara: Asil Yayıncılık.
- Kan, A. (2008). Ölçme aracı geliştirme. In S. Tekindal (Ed.), *Eğitimde ölçme ve değerlendirme* (pp. 247–287). Ankara: Pegem A Yayınları.
- Käpylä, M., Heikkinen, J. P., & Asunta, T. (2009). The influence of content knowledge on pedagogical content knowledge: The case of teaching photosynthesis and plant growth. *International Journal of Science Education*, 31(10), 1395–1415.
- Karaman, P. (2014). *Öğretmen adaylarının ölçme-değerlendirme okuryazarlıklarının belirlenmesi ve mikro-öğretim yoluyla geliştirilmesi*. Doktora tezi, On Sekiz Mart Üniversitesi Eğitim Bilimleri Enstitüsü, Çanakkale.
- Karaman, P., & Şahin, Ç. (2014). Öğretmen adaylarının ölçme değerlendirme okuryazarlıklarının belirlenmesi. *Abi Evran Üniversitesi Krşebir Eğitim Fakültesi Dergisi*, 15(2), 175–189.
- King, J. D. (2010). *Criterion-referenced assessment literacy of educators*. Unpublished doctoral thesis, University of Southern Mississippi, Hattiesburg, MS, USA.
- Lederman, N. G., & Gess-Newsome, J. (1992). Do subject matter knowledge, pedagogical knowledge, and pedagogical content knowledge constitute the ideal gas law of science teaching? *Journal of Science Teacher Education*, 3(1), 16–20.
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20–32.
- McMillan, J. H. (2014). *Classroom assessment: Principles and practice for effective standards-based instruction*. Boston, MA: Pearson Education.
- Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, 120, 285–296.
- Mertler, C. A. (2003). Pre-service versus in-service teachers' assessment literacy: Does classroom experience make a difference? *Annual Meeting of the Mid-Western Educational Research Association*, October 15–18, 2003, Columbus, OH, USA.

- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools, 12*(2), 101–113.
- Mertler, C. A., & Campbell, C. (2005). Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory. *Annual Meeting of the American Educational Research Association*, April 11–15, 2005, Montréal, Quebec, Canada.
- Ogan-Bekirođlu, F., & Suzuk, E. (2014). Pre-service teachers' assessment literacy and its implementation into practice. *The Curriculum Journal, 25*(3), 344–371.
- Özdamar, K. (1999). *Paket program ile istatistiksel veri analizi*. Eskişehir: Kaan Kitapevi.
- Özsevgeç, T., Çepni, S. ve Demirciođlu, G. (2004). Fen bilgisi öğretmenlerin ölçme-değerlendirme okur-yazarlık düzeyleri. *VI. Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresinde sunulmuş bildiri*, 16–18 Eylül 2004, İstanbul.
- Pellegrino, J. W., & Hilton, M. L. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, D.C.: The National Academies Press.
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator, 46*(4), 265–273.
- Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator, 36*(2), 115–131.
- Richardson, V. (1995). The consideration of belief in staff development. In V. Richardson (Ed.), *A theory of teacher change and the practice of staff development: A case in reading instruction* (pp. 102–119). New York, NY: Teachers College Press.
- Russell, M. K., & Airasian, P. W. (2012). *Classroom assessment: Concepts and applications*. New York, NY: McGraw-Hill.
- Shepard, L., Hammerness, K., Darling-Hammond, L., Rust, F., Snowden, J. B., Gordon, E., ... Pacheco, A. (2005). Assessment. In L. Darling-Hammond, & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 275–326). New York, NY: Jossey-Bass.
- Stacey, K., Helme, S., Steinle, V., Baturo, A., Irwin, K., & Bana, J. (2001). Preservice teachers' knowledge of difficulties in decimal numeration. *Journal of Mathematics Teacher Education, 4*(3), 205–225.
- Stiggins R. J. (1991). Assessment literacy. *Pbi Delta Kappan, 72*(7), 534–539.
- Stiggins, R. J. (2008). *Assessment manifesto: A call for the development of balanced assessment systems*. Portland, OR: ETS Assessment Training Institute.
- Stiggins, R. J. (2010). Essential formative assessment competencies for teachers and school leaders. In H. L. Andrade, & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 233–250). New York, NY: Routledge.
- Şahin, Ç., & Karaman, P. (2013). Sınıf öğretmenleri adaylarının ölçme ve değerlendirmeye ilişkin inançları. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 28*(2), 394–407.
- Tao, N. (2014). *Development and validation of Classroom Assessment Literacy Scales: English as a foreign language (EFL) instructors in a Cambodian higher education setting*. Doctoral dissertation, Victoria University, College of Education, Melbourne, Australia.
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research*. Thousand Oaks, CA: Sage.
- Tsagari, D. (2008). Assessment literacy of EFL teachers in Greece: Current trends and future prospects. *5th Annual EALTA Conference*, May 8–11, 2008 Athens, Greece.
- Türnüklü, E. B. (2005). Matematik öğretmen adaylarının pedagojik alan bilgileri ile matematiksel alan bilgileri arasındaki ilişki. *Eğitim Araştırmaları Dergisi, 21*, 234–247.
- Uluman, M. (2009). *Sınıf içi durum belirleme tekniklerine dayalı öğretimin öğrencilerin ölçme ve değerlendirme dersine ilişkin yeterlikleri üzerindeki etkilerinin incelenmesi*. Yüksek lisans tezi, Abant İzzet Baysal Üniversitesi Sosyal Bilimler Enstitüsü, Bolu.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education, 30*(3), 749–770.
- Walters, F. S. (2010). Cultivating assessment literacy: Standards evaluation through language-test specification reverse engineering. *Language Assessment Quarterly, 7*(4), 317–342.
- Wayman, J. C., Midgley, S., & Stringfield, S. (2006). Leadership for data-based decision-making: Collaborative data teams. In A. B. Danzig, K. M. Borman, B. A. Jones, & W. F. Wright (Eds.), *New models of professional development for learner centered leadership* (pp. 189–206). Mahwah, NJ: Erlbaum.
- William, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education, 11*, 49–65.
- Wise, S. L., Lukin, L., & Roos, L. L. (1991). Teacher beliefs about training in testing and measurement. *Journal of Teacher Education, 42*, 37–42.
- Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of Chinese college EFL teacher's experience. *TESOL Quarterly, 43*(3), 493–513.
- Yamtim, V., & Wongwanich, S. (2014) Study of classroom assessment literacy of primary school teachers. *Social and Behavioral Sciences, 116*, 2998–3004.
- Yıldırım, A., & Şimşek, H. (2008). *Sosyal bilimlerde nitel araştırma yöntemleri*. Ankara: Seçkin Yayıncılık.
- Zhang, Z., & Burry-Stock, J. A. (1997). Assessment practices inventory: A multivariate analysis of teachers perceived assessment competency. *Annual Meeting of the National Council on Measurement in Education*, March 25–27, 1997, Chicago, IL, USA.