Marmara Üniversitesi Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi • Temmuz 2023, 58: 166-185• ISSN: 2147-5202
DOI: 10.15285/maruaebd.1294902

ARAŞTIRMA MAKALESİ / RESEARCH ARTICLE

The Effectiveness of a Classroom Assessment Technique (CAT) in Measurement and Evaluation Classroom and Perceptions of Experimental Group towards CAT

Sınıf İçi Durum Belirleme Tekniklerinin (SİDBT) Ölçme ve Değerlendirme Dersi Üzerindeki Etkililiği ve Deney Grubunun SİDBT Yönelik Algıları

Müge ULUMAN MERT[1]

Abstract

Formative assessment such as classroom assessment techniques are used very little in higher education. When the literature is examined, it is seen that the use of these techniques makes significant contributions to the teaching process. The aim of this study is to find out the impact of instruction with classroom assessment techniques on students' level of competency in reliability and validity and the metaphors regarding these techniques. As a result of the study, it can be argued that teaching with the aid of classroom assessment techniques were effectual in increasing the students' competencies about reliability and validity. Parallel to this, it can be stated that most of the metephores produced by the experimental group are positive.

Keywords: Classroom assessment techniques, Formative assessment, Metaphors towards classroom assessment techniques.

Öz

Sınıf içi durum belirleme teknikleri gibi biçimlendirici değerlendirme, yükseköğretimde çok az kullanılmaktadır. Literatür incelendiğinde bu tekniklerin kullanımının öğretim sürecine önemli katkılar sağladığı görülmektedir. Bu çalışmanın amacı, Sınıf içi durum belirleme teknikleri ile öğretimin öğrencilerin güvenirlik ve geçerlik konusundaki yeterlilik düzeylerine etkisini ve bu tekniklere ilişkin metaforları ortaya çıkarmaktır. Araştırma sonucunda Sınıf içi durum belirleme teknikleri yardımıyla öğretimin öğrencilerin güvenirlik ve geçerlik konusundaki yeterliliklerini

1 Dr. Öğr. Üyesi, Marmara Üniversitesi Atatürk Eğitim Fakültesi Eğitim Bilimleri Bölümü, E-posta: mugeulumann@gmail.com, Orcid ID: 0000-0003-4155-3114

Makale Gönderim Tarihi: 09.05.2023 **Makale Kabul Edilme Tarihi:** 03.07.2023

artırmada etkili olduğu söylenebilir. Buna paralel olarak deney grubu tarafından üretilen metaforların çoğunun pozitif olduğu ifade edilebilir.

Anahtar Kelimeler: Sınıf içi durum belirleme teknikleri, Biçimlendirici değerlendirme, Sınıf içi durum belirleme tekniklerine yönelik metaforlar.

Geniş Özet

Giriş

Öğretmenlik mesleğinin oldukça önemli genel yeterliliklerinden biri alan eğitimiyle doğrudan ilgili olan ve öğretmenlerin sınıfta geçirdiği zamanın yadsınamaz bir kısmını oluşturan ölçme ve değerlendirmedir (Plake, 1993). Nitelikli bir öğretmenin, eğitim sürecinin sağlıklı yürütülebilmesi açısından ölçme ve değerlendirme yetenek ve yeterliliklerine sahip olması gerekmektedir (Anderson, 2008; Çakan, 2004; Darling Hammond, Wei ve Johnson, 2009; Stiggns, 2005; Stronge, Tucker ve Hindman, 2004; Welsh ve D'Agostino 2009). Alan yazında öğretmen adaylarının ve öğretmenlerin ölçme ve değerlendirme dersine ilişkin düşük okuryazarlık düzeylerine (Birgin 2007; Gül 2012; Karaman ve Şahin, 2014; Karaman ve Şahin, 2017), tatmin edici olmayan düzeyde de bilgi, becerileri ve yeterliliklere sahip olduklarını ortaya koyan çalışmalar bulunmaktadır (Çakan, 2004; Çalışkan, Uymaz ve Tekin, 2013; Demir Atalay, 2017; Evin Gencel ve Özbaşı, 2013; Gelbal ve Kelecioğlu 2007; Kilmen, Akın Köstereliođlu ve Köstereliođlu, 2007; Pektaş, 2010; Sabancı ve Yazıcı, 2017; Yaralı, 2017; Yaman ve Karamustaoğlu, 2011). Öğretmenlerin özellikle ölçme araçlarının geçerlik ve güvenirliklerini belirleme noktasında sorunlarla karşılaştıkları ve yeterli bilgi birikimlerinin olmadığı görülmektedir (Birgin ve Gürbüz, 2008; Çakan, 2004; Güven, 2001).

ABD'de 1990'lı yıllarda özellikle üniversitede verilen öğretim ve buna paralel olarak da öğrencilerin öğrenmelerinin niteliği bağlamında karşılaşılan sorunların çözümüne yönelik araştırmalar gerçekleştirilmiştir. Bu araştırmalar arasında öne çıkan çözüm önerilerinden birisi de Angelo ve Cross (1993) tarafından ortaya koyulan sınıf içi durum belirleme teknikleridir. Sınıf içi durum belirme teknikleri farklı birçok disiplinde işe koşulabilen, bilişsel öğrenme teorisiyle bağlantılı olan, öğretim süreci boyunca sistematik olarak uygulandığında öğretimin ve öğrenmenin niteliğini arttıracak toplam 50 teknik olarak tanımlanabilir (Angelo ve Cross, 1993).

Alan yazında sınıf içi durum belirleme teknikleri kullanılarak gerçekleştirilen lisans düzeyi farklı disiplinlere ait bir çok çalışma yer almaktadır. (Byon, 2005; Carduner 2002; Cross ve Palese, 2015; Nartgün ve Uluman, 2009; McNair 2000). Tüm bu çalışmalarda sınıf içi durum belirleme tekniklerinin öğrenme ortamını aktif hale getirdiği, ilgili konuya ilişkin öğrenci katılımını ve ilgisini arttırdığı, öğretim elemanına kendini geliştirmesi ve öğrencilerinin öğrenmelerini daha detaylı olarak takip edebilme imkânı sunduğu, son olarak da öğrencilere öğrenme ortamından keyif almalarına katkı sağladığı ortaya koyulmuştur. Bu bağlamda bu araştırmada sınıf içi durum belirleme tekniklerinin kullanımıyla gerçekleştirilen öğretimin, öğrencilerin ölçme ve değerlendirme dersi, güvenirlik ve

geçerlilik konularındaki yeterlilik düzeylerine olan etkisinin incelenmesi ve bu tekniklere ilişkin metaforların belirlenmesi amaçlanmıştır.

Yöntem

Araştırmanın amaçları doğrultusunda karma araştırma desenlerinden sıralı-açıklayıcı tasarım kullanılmıştır. Bu tasarımda öncelikle nicel daha sonra nitel verileri toplanır ve nitel verilerden nicel verileri doğrulamak, arttırmak ve açıklamak için faydalanılır (Creswell, Plano-Clark, Gutmann ve Hanson, 2003; Creswell, 2009). Araştırmanın nicel aşamasında öntest-sontest kontrol gruplu deneysel desen kullanılmış; nitel aşamasında ise bireysel tecrübelere dayanan olgu bilim (fenomenoloji) yaklaşımı kullanılmıştır. Araştırmanın çalışma grubunu bir devlet üniversitesinin resim öğretmenliği ana bilim dalında öğrenim gören, deney grubunda 34 kontrol grubunda ise 33 öğrenci olmak üzere toplam 67 öğrenciden oluşmaktadır. Araştırma kapsamında araştırmacı tarafından geliştirilen iki farklı araç kullanılmıştır. Araçlardan ilki deney ve kontrol grubunda yer alan öğrencilerin ölçme ve değerlendirme dersi güvenirlik ve geçerlik konularına ilişkin başarılarını ölçmek amacıyla geliştirilen çoktan seçmeli test, ikincisi ise öğrencilerin metaforlarını belirlemek için geliştirilmiş yazılı formdur. Elde edilen nicel verilerin çözümlenmesinde, tek faktör üzerinde tekrarlı ölçümler için ANOVA kullanılmış; nitel verilerin çözümlenmesinde ise içerik analizinden faydalanılmıştır.

Bulgular

Öğretmen adaylarının ölçme ve değerlendirme dersinde, sınıf içi durum belirleme teknikleri kullanılarak gerçekleştirilen, güvenirlik ve geçerlilik konuları öğretiminin, öğrencilerin başarılarına etkilerinin ve bu tekniklere yönelik deney grubu öğretmen adaylarının algılarının araştırıldığı çalışmadan elde edilen sonuçlar bu bölümde açıklanmıştır.

Araştırmanın birinci alt amacı doğrultusunda, deney ve kontrol grubu öğretmen adaylarının ön test ve son test toplam puanları arasında farkın manidar olduğu görülmektedir. Bu bulgu, iki farklı grupta yer alan öğrencilerin farklı ölçme durumlarından aldıkları geçerlik ve güvenirlik testi puanlarının farklılaştığını göstermektedir. Bununla birlikte, öğretmen adaylarının ön test son test puanları arasında manidar bir farklılık vardır. Buna bulguya dayanarak, grup ayrımı yapılmaksızın öğretmen adaylarının geçerlik ve güvenirlik testi puanlarının ölçümlere göre manidar bir şekilde farklılaştığı belirtilebilir. Farklı işlem gruplarında yer alma ile farklı zamanlarda ölçümü alınan faktörler dikkate alındığında öğretmen adaylarının geçerlik ve güvenirlik testi puanları üzerindeki ortak etkisinin manidar olduğu görülmektedir. Bu bağlamda sınıf içi durum belirleme teknikleri kullanılarak öğretim yapılan deney grubu öğretmen adaylarının puanlarında deney öncesine göre gözlenen değişme, kontrol grubundaki öğretmen adaylarının puanlarında gözlenen değişmeden farklıdır. Yani deney ve kontrol grubu öğretmen adaylarının puanları deney grubuna uygulanan deneysel işleme bağlı olarak farklılık göstermektedir. Başka bir değişle deney ve kontrol grubu öğretmen adaylarının puanlarında gözlenen bu farklılığın, sınıf içi durum belirleme teknikleriyle işlenen öğretimden kaynaklandığı ifade edilebilir.

Araştırmanın ikinci alt amacı doğrultusunda, deney grubu öğretmen adaylarının sınıf içi durum belirleme tekniklerin ilişkin toplam 34 geçerli metafor ürettikleri ve toplam 27 farklı metafor olduğu görülmektedir. En fazla tekrarlanan metafor ayna (3) olurken bu metaforu 2 tekrar sayısıyla lunapark, yol, fırça, fotoğraf ve tiner izlemektedir. Diğer tüm metaforlar ise birer kez yazılmıştır. Bu doğrultuda deney grubu öğretmen adaylarının sınıf içi durum belirleme tekniklerin ilişkin çok sayıda farklı metafor ortaya koyduğu belirtilebilir. Deney grubu öğretmen adaylarının sınıf içi durum belirleme tekniklerine ilişkin ürettikleri metaforların nedenleri arasında en fazla 'Yansıtıcıdır' ve 'Destekleyicidir' in tekrarlandığı belirtilebilir. Bu iki nedeni yine olumlu olarak değerlendirilebilecek 'Yönlendiricidir' ve 'Amaca götüren araçtır' nedenleri takip etmektedir. Olumsuz olarak ele alınabilecek nedenler ise üçer metaforla bağlantılı olan 'Yorucudur' ve 'Faydalı değildir' nedenleridir. Geçerli metaforların temalara atanması metaforlar ve bağlantılı oldukları nedenlerin tekrar incelenmesi üzerine gerçekleştirilmiştir. Ortaya çıkan beş tema; eğlenceli olması, öğrenme sürecini desteklemesi ve düzenlemesi, zaman alıcı ve yorucu olması, öğrenmeyi ve sürecini yansıtması, olmasa da olur durumu şeklinde sıralanabilir.

Tartışma

Deney ve kontrol grubu öğretmen adaylarının, öğretim sürecinin başında testten elde edilen ortalamalar dikkate alındığında güvenirlik ve geçerlilik konularındaki yeterliliklerinin neredeyse aynı düzeyde olduğu; öğretim sürecinin sonunda ise her iki grubun da yeterlilik düzeylerinde bir artışın yaşandığı görülmüştür. Bununla birlikte, deney grubunda gözlenen artışın kontrol grubuna nazaran daha yüksek olduğu ve bu farkın manidar olduğunu sonucuna ulaşılmıştır. Bu bağlamda sınıf içi durum belirleme tekniklerinden faydalanılarak gerçekleştirilen öğretimin, öğretmen adaylarının güvenirlik ve geçerlilik konularındaki yeterlilik düzeylerini arttırmada etkili olduğu belirtilebilir. Elde edilen bu sonuç, literatürde yer alan farklı disiplin ve düzeylerde gerçekleştirilmiş çalışmalara ait sonuçlarla (Carduner 2002; Eckert et al. 1997; Murdock, 2018; Walker 1991; Schwarm and VanDeGrift, 2002; Byon, 2005; Nartgün ve Uluman, 2009; Nartgün, 2010; Gaeddert 2003) paralellik taşımakla birlikte sınıf içi durum belirleme tekniklerinin etkililiğine ilişkin kanıt niteliği taşımaktadır.

Her iki tema ve içerdikleri metaforlar dikkate alındığında, sınıf içi durum belirleme tekniklerinin öğretim sürecini daha nitelikli hale getirdiği, öğretmen adaylarının gelişimlerini gözlemleme fırsatı sunduğu ve öğrenmenin gerçekleşebilmesi için rehber olduğu sonuçlarına ulaşılabilir. Deney grubu öğretmen adaylarının sınıf içi durum belirleme tekniklerine ilişkin ortaya koyduğu bu olumlu sonuçlar literatürde yer alan birçok çalışma (Gaeddert, 2003; Goldstein, 2007; Beard, 1993; Cross ve Palese, 2015; Soetaert, 1998) sonuçlarını destekler niteliktedir.

Bu sonuçlara dayanılarak ölçme ve değerlendirme dersi güvenirlik ve geçerlilik konularının öğretiminde sınıf içi durum belirleme tekniklerinin öğretim süreçlerine dâhil edilmesi önerilmektedir. Bununla birlikte özellikle daha düşük okuryazarlık düzeyine ve tatmin edici olmayan düzeyde yeterliliğe sahip olunan ölçme ve değerlendirme dersinin tüm konularında sınıf içi durum belirleme teknikleri kullanımının öğretim sürecine katkısının incelenmesi önemli görülmektedir. Ayrıca Türkiye'de gerçekleştirilen çalışmaların sınırlılığı göz önüne alındığında farklı disiplin ve düzeylerde

sınıf içi durum belirleme tekniklerinin öğretim sürecine katkılarının incelenmesinin faydalı olacağı düşünülmektedir.

Introduction

The sum of the knowledge, skills and attitudes necessary for teachers to perform their profession effectively and efficiently is called teachers' competencies. They are the key qualifications for teachers if they want to maintain their personal development, increase student success and offer highly-qualified education. From this perspective, it is an unavoidable task for teachers to improve their skills and competencies in order to keep up with the professional requirements of the job as the world is becoming more developed (MEB, 2017).

One of the most important overall competencies that teachers need to have is measurement and evaluation, which is directly connected with subject teaching and occupies a considerable part of teachers' classroom time (Plake, 1993). A highly-qualified teacher must possess measurement and evaluation skills and competencies in order to carry out the education process properly (Çakan, 2004; Stronge et al., 2004; Stiggns, 2005; Darling-Hammond et al., 2009; Welsh & D'Agostino, 2009). A teacher equipped with these competencies will naturally contribute to achieving the objectives of education and thus increasing student success, and assessment of the methods, instruments and materials used (Koh 2011; Popham, 2002; Shepard, 2000; Stiggins, 2005; Xu et al., 2009). At this point, education faculties seem to have a big part to play in teaching measurement and evaluation skills and competencies to the teachers of the future.

In a narrow scale, the extent of candidate teachers' possessing these qualifications is measured by academicians teaching the lessons in Turkey. Later, these qualifications are checked with the Test on Educational Sciences as a part of the nation-wide Public Personnel Selection Examination (KPSS) in macro scale. However, no information is available about achievement levels of the prospective teachers in either of the measurements above. Nevertheless, the literature shows that teachers and prospective teachers have only little literacy of measurement and evaluation (Birgin et al., 2009; Gül 2012; Karaman et al., 2014; Karaman et al., 2017) and unsatisfactory levels of knowledge, skills and competencies (Çakan, 2004; Çalışkan, et al., 2013; Demir Atalay, 2017; Gencel et al., 2013; Gelbal et al., 2007; Kilmen, et al., 2007; Pektaş, 2010; Sabancı et al., 2017; Yaralı, 2017; Yaman et al., 2011). It has been reported that teachers face difficulties particularly in checking the reliability and validity of measurement and evaluation instruments, and they lack adequate knowledge to do so (Birgin et al., 2008; Çakan, 2004; Güven, 2001). It can be argued that candidate teachers attribute both current and future challenges regarding measurement and evaluation competencies to the unsatisfactory undergraduate education they receive (Sabancı et al., 2017).

In the 1990s, researches were conducted in the USA in order to solve the problems encountered in particular context of university education and the consequential quality of students' learning. As a result of such attempts, classroom assessment techniques emerged as a prominent solution proposal, among others, put forward by Angelo et al., (1993). Classroom assessment techniques can be defined as a total of 50 techniques that can be employed in many different disciplines and are linked to the

cognitive learning theory, and that will improve the quality of teaching and learning when applied systematically throughout the teaching process (Angelo et al., 1993). Using these techniques provides more formative data than assessing the outcome of classroom success and failure (Beard, 1993). They can be used for a myriad of purposes ranging from assessment of readiness to the level of gaining critical thinking skills. Thanks to these techniques, instructors can assess the students' learning and comprehension informally. By the same token, students can reflect on their own learning before taking high-stake tests. Also, as an indirect benefit, the feedback obtained in this way serves to give instructors feedback regarding pedagogical and andragogic strategies and ways to improve themselves for attaining better learning outcomes (Beard, 1993; Cross et al., 2015).

Classroom assessment techniques constitute an authentic assessment tool owing to two qualities. First, they encourage both students and instructors to learn constructively in an interactive classroom setting. Second, unlike other assessment techniques, they offer feedback on the effectiveness of instruction while encouraging learning of the topic or subject (Angelo et al., 1993).

The techniques have six distinct characteristics and seven basic assumptions. But it should be remembered that the techniques can be used successfully only if the classroom assessment cycle is employed which consists of three stages with three sub-stages under each since classroom assessment techniques are regarded as a project (Angelo et al., 1993). The literature provides a lot of studies using these techniques in various disciplines serving to various purposes. Examples of studies carried out in undergraduate courses with classroom assessment techniques include foreign language education (Carduner, 2002); sociology (Eckert et al. 1997); introduction to psychology (Walker 1991); mathematics (Murdock, 2018); introduction to using computer (Schwarm et al., 2002); Korean as a foreign language (Byon, 2005); instructional principles and methods (Eminoğu Küçüktepe, 2015) and measurement and evaluation course (Nartgün et al., 2009). As for the postgraduate level, there are studies reporting the use of these techniques in distant education of courses such as technology education (Gaeddert, 2003), early childhood teaching education (McNair, 2000), and research techniques (Nartgün, 2010; Cross et al., 2015). The previous studies have revealed that classroom assessment techniques revive the in-class learning environment, increase student interest and participation, help instructors develop themselves and monitor the students' learning more closely, and allow the students to enjoy the learning environment. On the other hand, there are studies which prove the opposite (Cottell et al., 1998; Simpson Beck, 2011).

Departing from what is mentioned above, it can be said that instructors often assess their students' learning with an outcome-oriented approach by grading a single test or assignment, but they do not apply the classroom assessment techniques in the higher education context although they stand as a type of formative assessment reflecting learners' improvement (Goldstein, 2007). It seems necessary to inspire the use of these techniques to elevate the quality of learning in tertiary education and thus mitigate the resulting problems while adding new examples to the existing small number of such studies in Turkey at the same time. From the perspective of education faculties, it seems significant to teach measurement and evaluation competencies, which are particularly perceived to be hard and complicated by candidate teachers (Gronlund, 2002), so that the faculty students can build a highly-qualified education process in the future. In this regard, it is known that teachers face hardship and

suffering from lack of knowledge in identifying reliability and validity of measurement instruments (Birgin et al., 2008; Çakan, 2004; Güven, 2001); therefore, candidate teachers should be trained on methods and techniques other than those currently in use.

On the grounds of the above, the aim of this study is to find out the impact of instruction with classroom assessment techniques on students' level of competency in reliability and validity and the metaphors regarding these techniques. Under the main aim of the study, answer is sought to two questions as follows:

1. Is there a significant difference between success levels of the experiment group (candidate teachers learning with the aid of classroom assessment techniques) and the control group (those not using the techniques) in reliability and validity topic of measurement and evaluation?
2. What metaphors are used by the experiment group in referring to their perceptions of classroom assessment techniques?

Methodology

Research Model

This study was carried out with sequential descriptive pattern, a blend of qualitative and quantitative methods, due to the study objectives. In other words, a mixed method was used in the research. As the first step, the experiment and control group were formed through random assignment and effectiveness of classroom assessment techniques was tested by looking at the pre and post test scores of the students in both groups. For this reason, pre-test post-test control group experimental design was used for the quantitative aspect. As required by the other goal of the study, to determine the metaphors used by the experimental group about their perceptions of the classroom assessment techniques, phenomenology was preferred as a qualitative research approach since it is based on individual experiences. In this approach, there is no single reality and reality is based on personal perceptions (Giorgi et al., 2003).

Study Group

In the study, 67 undergraduate students enrolled in the spring term of 2018-2019, arts teaching program in a Turkish state university took part. The students were divided into two groups with simple random sampling method, appointing the participants into one experimental group and one control group, respectively on a random basis. There were 34 students in the experimental group and 33 in the control group.

Data Collection Tools and Procedure

In this study, two different instruments were used to collect data. One of them is a multiple-choice test developed by the researcher in order to measure the success of both groups in reliability and validity as a requirement of measurement and evaluation course. This instrument was developed by following the test development stages (Atılgan et al., 2007; Baykul, 2015). In first place, a table of specifications was prepared by considering the purpose of the test. It was planned to prepare the test

by including 30 items 15 of which covering reliability and another 15 covering validity according to the table. A total of 64 indicative items were drafted and sent to three independent experts of measurement and evaluation. The test items were reviewed by following the expert view and then a pilot test form was drawn up. The pilot form was applied to 240 undergraduate students, including the study participants, in a public university who took and passed measurement and evaluation course during the previous semester. The responses were graded as 0 or 1 so that test and item analyses could be performed on the data sets. The test and item statistics were conducted on the data set by using Excel, and the final items in the test were selected as a result. The test finally consisted of 30 items corresponding to the top score of 100. The item statistics of the test can be seen in Table 1.

Table 1.
Multiple-Choice Test Item Statistics

Item	p	Q		s_j	r_{jx}	r_j
1	0,52	0,48	0,25	0,50	0,38	0,19
2	0,74	0,26	0,19	0,44	0,34	0,15
3	0,56	0,44	0,25	0,50	0,53	0,26
4	0,34	0,66	0,22	0,47	0,48	0,23
5	0,51	0,49	0,25	0,50	0,67	0,33
6	0,56	0,44	0,25	0,50	0,5	0,25
7	0,64	0,36	0,23	0,48	0,63	0,30
8	0,47	0,53	0,25	0,50	0,44	0,22
9	0,68	0,32	0,22	0,47	0,53	0,25
10	0,37	0,63	0,23	0,48	0,56	0,27
11	0,42	0,58	0,24	0,49	0,49	0,24
12	0,59	0,41	0,24	0,49	0,62	0,30
13	0,42	0,58	0,24	0,49	0,41	0,20
14	0,52	0,48	0,25	0,50	0,55	0,27
15	0,72	0,28	0,20	0,45	0,66	0,30
16	0,4	0,60	0,24	0,49	0,61	0,30
17	0,45	0,55	0,25	0,50	0,72	0,36
18	0,61	0,39	0,24	0,49	0,46	0,22
19	0,28	0,72	0,20	0,45	0,33	0,15
20	0,49	0,51	0,25	0,50	0,45	0,22
21	0,63	0,37	0,23	0,48	0,59	0,28
22	0,56	0,44	0,25	0,50	0,65	0,32
23	0,75	0,25	0,19	0,43	0,4	0,17
24	0,3	0,70	0,21	0,46	0,47	0,22
25	0,44	0,56	0,25	0,50	0,59	0,29
26	0,51	0,49	0,25	0,50	0,73	0,36
27	0,57	0,43	0,25	0,50	0,46	0,23
28	0,55	0,45	0,25	0,50	0,53	0,26
29	0,61	0,39	0,24	0,49	0,57	0,28
30	0,53	0,47	0,25	0,50	0,6	0,30

As seen in Table 1, item difficulty indices were found to be between 0,28 and 0,75. There were two difficult items (item no 19 and 24) and three easy items (item no 2, 15, and 23), leaving the rest in the range of medium difficulty. It implies that the test as a whole reached medium difficulty level, which is the optimum level of difficulty for achievement tests. The statistics also showed that item discriminatory indices were above 0,30 varying in the range of 0,33 and 0,73. By using the values in the table, the KR-20 coefficient was calculated which particularly indicates internal reliability of achievement tests graded as 0 or 1, and the reliability was found as 0,91. These results proved that this data collection instrument met the criterion of reliability for use.

The other instrument of data collection was developed in order to find out the metaphors used by the respondents. As the first step, the literature was reviewed for studies determining metaphors (Duman et al., 2014; Gök et al., 2012; Günal, 2014; Güvendir et al., 2016; Taşgın et al., 2015; Tunç et al., 2018). Then, a written form was prepared accordingly which included the statement 'Classroom assessment techniques are like …, because…'.

Before starting, the units of reliability and validity to be taught in this study were determined and the instruction was planned for both the experimental and control groups. When necessary preparations were completed, the implementation was started in week 3 during the 2018-2019 academic year spring semester and it was completed by week 6. Prior to the onset of the practical stage, pre-tests were given to both groups during class hours to unearth their preliminary competency in reliability and validity. Apart from that, the students were informed about the content of the course, weekly syllabuses, methods and techniques to be used, and the references at the beginning of the semester. The teaching syllabus was provided for both of the groups. No explanations were made to the control group as it was planned to teach the content as it used to be by then. On the contrary, the experimental group was trained about classroom assessment techniques along with example uses in a variety of disciplines. It was made obligatory for this group to participate in the applied training with a minimum quote of absence. Below is given a summary definition and examples of the techniques to be used by the experimental group.

Misconception/Preconception Check: The biggest obstacle before learning new things is the wrong or incomplete knowledge of students about the topic, rather than the missing parts in their preliminary knowledge. The wrong or incomplete learning is likely to lead to misconceptions and prejudices; therefore, it would be useful to understand the extent they could reach. The focus of this technique is on exposing knowledge and beliefs that could hinder or block learning. The instruction goals of this technique include learning concepts and theories as well as methods and realities on any topic, and explicating and developing newly-learnt information.

Approximate Analogies: This technique helps measure how well the students could comprehend crucial points in the overall class, reading passages or the monitoring material used as teaching materials. It also boosts recalling of the main parts of the class and understanding of those parts by students and infusing accurate information to students. The technique is most useful when used in complex classroom settings which have a regular flow with too much content, phenomena and rules. In this technique, the students are given a fully or partially blank template to fill in the blanks on it

within a limited period of time. Approximate analogies can be set as an assignment or an in-class activity. The technique has instructional goals such as arousing interest in the lesson, increasing the learners' concentration, listening skills, proper studying skills and habits, and learning the terms and concepts concerning the topic.

The Muddiest Point: In spite of being one of the easiest classroom assessment techniques, *the muddiest point* is quite efficient because the time and energy commitment is minimum, but it is rewarded with quite a large amount of learning. The students are asked to rapidly provide a written answer to the question "What was the muddiest point in …?" so that feedback can be obtained about the aspect which is found the most confusing or ambiguous by the students. With the help of the feedback, the lecturer reflects on the most difficult things to learn from the learners' perspective, the subtopics which deserve more attention, and time management regarding these subtopics. The question concerning the muddies point must be addressed at the very end of a class, a debate or presentation, or right after a reading assignment. The best setting to use this technique could be the classes with intensity of new information to present before starting the class because those classes probably have the highest amount of "the muddiest point". The instructional goals are to develop suitable studying skills, strategies and habits, concentration and listening skills of students, and teaching concepts, theories and concrete facts about the topic in question.

Applications Cards: In this technique; once the learners have an idea about an important principle, generalization, theory or method, the teacher distributes small cards to the students and tells them to write down at least one example which embodies what they have just learnt and which is associated with real life at the same time. These cards provide insight into the extent at which students are able to put their learning into practice for the lecturers. Overall, the technique allows learners to brainstorm about ideal examples of application and associate new learning with their previous experience. In this way, the students can realize that the principles and theories learnt in class correspond to things in real life, and this would boost the learners' interest and motivation for learning. As an additional benefit, the technique makes it possible for the students to learn from peers' examples more than the examples presented by the lecturer or course books. As a result, peer learning is activated and the lecturer can benefit from a new and renewable source of examples. As for the instructional goals of this technique, they can be listed as the ability to make plausible deductions from observations, enhancing suitable studying skills, strategies and habits and self-reflection capability, and lastly learning theories and concepts concerning the topic in question.

After the overall procedure was completed, the pre-test which was previously given to the experimental and the control group was administered to the both groups as a post-test this time. Besides, the other data collection instrument was applied to the experimental group with the purpose of revealing the metaphors they used to refer to the classroom assessment techniques. The procedure was completed when the students filled out the form.

Data Analysis

As mentioned above, the first set of data was collected by applying pre-test post-test control group experimental design. The purpose of using that instrument was to find an answer to the first research

question of the study. For analysing the data in this model, one-way repeated measures ANOVA can be used to demonstrate the effectiveness of the process on the condition that the assumptions are fulfilled (Howitt et al., 1997; Münke, 1997 as cited by Büyüköztürk, 2001). In order to decide whether this statistical analysis is applicable for the current study data, the assumptions were tested at first.

In the analysis, the independent variable was reached through an interval scale, and independent observations were ensured. Moreover, the data were subjected to repeated measurements. The independent variable scores were seen to have a normal distribution across each category as evidenced in Table 2. According to the same table, the variances of the differences between all combinations of related groups must be equal.

Table 2.

Values regarding the Assumptions of One-Way Repeated Measures ANOVA

Test	Group	Kolmogorov-Smirnov Test			Skewness	Kurtosis	Levene Test			
		Statistics	sd	p			F	d1	d2	p
Pre-test	Exper.	0.138	34	0.102	0.953	2.009	0.00	1	65	0.987
	Control	0.146	33	0.072	0.951	2.109				
Post-test	Exper.	0.121	34	0.200	0.002	-1.099	0.521	1	65	0.473
	Control	0.155	33	0.044	0.643	-0.389				
Box's M = 0.993, p>0.05										

It can be seen in Table 2 that the significance values (p) for the Kolmogorov-Smirnov normality test were above 0.05, the threshold value, and there was normality convergence except for the control group post-test distribution. Yet, the significance value below the threshold was so close to the threshold. The other values indicating normal distribution are skewness and kurtosis. The former was found to be within the reference values (-1 – +1) for all subgroups, while kurtosis exceeded the limits except for the control group post-test distribution. However, taking into consideration that the values were not too distant from the limits and the graphics as the other n-normality indicators (histogram, Q-Q plot, P-P plot) prove convergence of normality in the subgroups and there were no extreme values in the distributions, it can be said that normality assumption was met for each of the groups. In addition to this, the variance tests stand strong against violation of the assumptions especially when the subgroups have equal or close numbers of participants (Field, 2009). Another finding was provided by Levene and Box's M tests in the table, which implies that covariances were equal for binary combinations of the variances and measurements. As a result, the assumptions were fulfilled and the data set was found to be suitable for using One-Way Repeated Measures ANOVA.

When it comes to the second research question, data analysis was performed with content analysis. Content analysis refers to surveying and categorizing verbal and written data for the ultimate aim of classifying and summarizing such data for a particular purpose, measuring specific concepts, and reaching certain conclusions accordingly (Tavşancıl et al., 2001).

In the study, the metaphors offered by the experimental group were analysed and interpreted at four steps: (1) naming, (2) sorting, (3) eliciting themes, (4) reliability and validity check (Saban, 2008; 2009).

In the naming step, a total of 34 metaphors were elicited along with their respective justifications, and they were analysed in alphabetical order. In the sorting step, the metaphors and justifications were analysed to find no invalid items. Next, the justifications provided for the metaphors were classified by taking into account the similarities among the metaphors. Then, the themes explaining the metaphors were elicited according to the common aspects of the metaphors. Finally, all the preceding phases were reported in detail and direct quotations were included in connection with the metaphors under each theme so as to prove the reliability and validity of the process. At the same time, a measurement and evaluation expert was consulted to find out whether the themes represent the metaphors concerned. In this scope, the expert was asked to send each metaphor to the relevant category. Then, the pairings made by the researcher and the measurement and evaluation expert were analysed with Miles and Huberman's (1994) formula (Reliability= number of agreement / number of agreement + disagreement X 100). In this way, the coefficient of concordance, the indicator of reliability, was calculated as 91% (31/31+3 X 100). The figure was above the threshold value of 90% and thus evidence of reliability.

Findings

First of all, descriptive statistics for both groups were examined. Table 3 shows the descriptive statistics of pre and post-test results of the experimental and control group students.

Table 3.
Descriptive Statistics

Tests	Groups	$\bar{x}$	ss	Lowest Score	Top Score	N
Pre-test	Experimental Group	26,44	8,20	9,99	53,28	34
	Control Group	25,83	8,16	9,99	49,95	33
Post-test	Experimental Group	70,52	13,13	49,95	93,24	34
	Control Group	57,12	14,99	29,97	89,91	33

It can be seen in Table 3 that the experimental group's pre-test scores ranged from 10 (9,99) to 53 (53,28), with the mean score of 26,44 and standard deviation of 8,20. Similarly, the control group obtained pre-test scores between 10 (9,99) and 50 (49,95), with mean score of 25,83 and standard deviation of 8,16. Although these figures show a proximity between the groups, the experimental group completed the study with a slightly higher average value. As for the post-test scores, they were found to be between 50 (49,95) and 93 (93,24) in the experimental group, whereas the values were ranging from 30 (29,97) to 90 (89,91) in the other group. The mean scores were 70,52 and 57,12 for the experimental and control group, respectively. Lastly, the groups recorded standard deviations of

13,13 and 14,99, respectively. It can be argued that the experimental group could obtain higher scores than the control group.

In order to determine the significance of the differences between the experimental group's and control group's pre and post-test scores, gruopxmeasurement formula was applied, and the effect of the results was exhibited in Table 4.

Table 4.
One-Way Repeated Measures ANOVA Results of Experimental/Control Group Pre and Post-Test Scores

Variance Source	Sum of Squares	sd	Mean Square	F	p	η2
Between Groups						
Group (E/C)	1644,534	1	1644,534	7,88	,007	,108
Error	13574,237	65	208,834			
Within Groups						
Measurement (Pre-Post)	47546,244	1	47546,244	847,13	,000	,929
Group*Measurement	1370,078	1	1370,078	24,41	,000	,273
Error	3648,228	65	56,127			

As seen in Table 4, there was a significant difference between the total pre and post-test scores of the experimental and control group students ($F_{(1,65)} = 7.88$, $p<.05$). This finding reveals that the participants in two different groups got different marks from the reliability and validity tests in different measurement situations.

There was found a significant difference between pre and post-test scores of the students ($F_{(1,65)} = 847.13$, $p<.05$). Thus, it can be said that the students' reliability and validity test scores differed significantly by measurement, regardless of the type of group.

According to Table 4, considering the inclusion in distinct groups and the factors measured at different times, there was a significant common effect on the students' scores in the reliability and validity tests ($F_{(1,65)} = 24.41$, $p<.05$). So the change in the experimental group's scores from pre to post-test was regarded to be distinctive compared to the change of scores in the control group. In other words, the experimental group could obtain higher scores as a result of the experimental procedure. In a nutshell, it can be claimed that the difference in favour of the experimental group was achieved via teaching through classroom assessment techniques.

As regards to the effect size obtained from the analysis (.108, .929, and .273, respectively), it can be suggested that the effect between the groups, measurement and Group*Measurement had a high level of effect.

In relation to research question two in the study, attempt was taken to find out what metaphors were used by the respondents in the experimental group to explain the classroom assessment techniques. To this end, the frequency values for the acceptable metaphors were listed in Table 5.

Table 5.
Accepted metaphors developed by experimental group participants concerning CATs and respective frequencies

Metaphor	f	Metaphor	f	Metaphorı	f
Amusement park	2	Angary	1	Drier	1
Mirror	3	Brush	2	Avocado	1
Beans	1	Easel	1	Pipet	1
Close friend	1	Model	1	Smart board	1
Road	2	GPRS	1	Concealer	1
Weekly	1	Eraser	1	Photograph	2
Car	1	Lamp	1	Bonus	1
Load	1	Music	1	Protein powder	1
Goggles	1	Map	1	Thinner	2

It is seen in Table 5 that the participants in the experimental group generated a total of 34 acceptable metaphors regarding CATs, 27 of them were distinct entries. The most repeated metaphor was noted as "mirror" (3), followed by "amusement park" (2), and "road", "brush", "photograph", and "thinner". The rest of the metaphors were mentioned only once each. It can thus be suggested that a wide assortment of metaphors regarding CATs were referred to by the experimental group. The frequency values regarding the justifications of the respective metaphors are given in Table 6.

Table 6.
Justifications and frequencies of the metaphors offered by the experimental group to explain CATs

Justification	f
Reflective	9
Supportive	5
Just instrumental	4
Guiding	4
A second chance to remedy the mistake or deficiency	4
Tiring	3
Unuseful	3
Amusing	2

As can be understood from Table 6, the highest number of repetitions was recorded under 'Reflective' and 'Supportive' as justifications for the metaphors. These items were followed by another couple of positive ideas as 'Guiding' and 'Just instrumental'. On the other hand, two justifications had negative connotations such as 'Tiring' and 'Unuseful' and each of them was connected with three metaphors.

Later, the metaphors and underlying justifications were re-examined and the valid metaphors were appointed to the themes deduced from the content analysis. A total of five study themes were found: "amusing", "supporting and organizing the learning process", "time-consuming and tiring",

"reflective of learning and the overall process of learning and dispensable status. The distribution of the metaphors by themes is given in Table 7.

Table 7.

Accepted metaphors suggested by the experimental group regarding CATs and corresponding themes

Theme	Metaphors	f	Number of Metaphors
Amusing	Amusement park (2)	2	1
Supporting and organizing the learning process	Close friend (1), Road (2), Car (1), Brush (2), Easel (1), Eraser (1), GPRS (1), Music (1), Thinner (2), Pipet (1), Concealer (1), Bonus (1), Protein powder (1), Map (1)	17	14
Time-consuming and tiring	Load (1), Angary (1), Avocado (1), Smart board (1),	4	4
Reflective of learning and the overall process of learning	Mirror (3), Weekly (1), Model (1), Lamp (1), Photograph (2), Goggles (1)	9	6
Dispensable status	Beans (1), Drier (1)	2	2

Table 7 demonstrates 27 different metaphors along with the related themes. In this table, it is seen that the biggest number of metaphors (17) were appointed to 'supporting and organizing the learning process' covering a total of 14 different metaphors. In the scope of this particular theme, it can be argued that the participants implied that CATs added to their learning processes. Below are direct citations from the respondents concerning the metaphors and justifications.

'Classroom assessment techniques are like an easel because they provide comfort for the teacher.'

'Classroom assessment techniques are like music because they help me understand the topics.'

'Classroom assessment techniques are like an eraser because they help me clear my wrong learning and relearn.'

The second most popular theme with the highest number of metaphors (6) and repetitions (9) was 'reflective of learning and the overall process of learning' again with positive connotations. Under this theme, it seems that the experimental group respondents placed emphasis on CATs in that they provide feedback regarding their learning and allow them to interpret this process. Below are direct quotations regarding this theme;

'Classroom assessment techniques are like a mirror because they show what I understand and what I don't understand.'

'Classroom assessment techniques are like goggles because they make me see myself better in the class.'

'Classroom assessment techniques are like a photograph because they show what is complete and what is incomplete.'

Unlike the other two, the theme with the third highest number of metaphors (4) and repetitions of metaphors (4) was 'time-consuming and tiring', which has a negative connotation. From this theme, it can be understood that the participants felt tired during the use of the CATs andthey found the techniques tedious, which was not anticipated due to the features of the techniques. This finding is supported with some direct quotations below.

'Classroom assessment techniques are like a burden because it is tiring to constantly do things.'

'Classroom assessment techniques are like angary because they keep you needlessly busy…'

'Classroom assessment techniques are like an avocado because they are tasteless despite being useful.'

Lastly, the remaining two themes had the same frequencies (2) but with dissimilar numbers of metaphors under each. Of these themes, a positive mindset is thought to underlie'amusing', whereas 'dispensable situation' seem to have a neutral ground. These themes are exemplified with the following citations:

'Classroom assessment techniques are like an amusement park because it is amusing and colourful to attend the activities.'

'Classroom assessment techniques are like a drier because it is fine with or without them...'

Conclusion and Discussion

This study was carried out to shed light onto the effect of instruction with classroom assessment techniques (CATs) on students' competencies in reliability and validity and to identify the metaphors used by the students in connection with CATs. First of all, the average scores in the pre-test reveal that the experimental and control group students were almost equally competent in reliability and validity of measurement and evaluation before the implementation of the teaching. However, at the end of the teaching, the competency level of both groups increased. But the increase in the experimental group was significantly higher. Therefore, it can be argued that teaching with the aid of classroom assessment techniques were effectual in increasing the students' competencies about reliability and validity. This finding is in compliance with other studies carried out in various disciplines and levels of education (Walker 1991; Eckert et al. 1997; Carduner 2002; Schwarm et al., 2002; Gaeddert 2003; Byon, 2005; Nartgün & Uluman, 2009; Nartgün, 2010; Murdock, 2018; Jawad, Majeed & ALRikabi, 2021). Also, it can be taken as evidence for the effect of CATs.

As for the descriptive metaphors, a total of 34 plausible items were derived along with the respective justifications. Taking into account the metaphors only, all of them have positive references except for two. But when considered together with the justifications, the number of negative metaphors was seen to increase to six. The 34 metaphors covering 27 distinct ideas were categorized under five themes. Among the others, the themes 'supporting and organizing the learning process' and 'reflective of learning and the overall process of learning' hold the highest number of metaphors. Considering the both themes and their coverage, it can be inferred that classroom assessment techniques could

better the instruction process, allowed monitoring the students' progress, and played a guiding role for realization of learning. These positive results concerning the experimental group's interpretation of classroom assessment techniques are in conformity with many other studies (Gaeddert, 2003; Goldstein, 2007; Beard, 1993; Cross et al., 2015; Soetaert, 1998). Furthermore, it was understood from the metaphors that the participants in this division find it fun to use CATs. Contrarily, the techniques were found exhaustive and time-consuming by some of the participants in addition to the techniques' not advancing their learning or not bringing in any positive effect. The literature also offers some studies which refute positive effects of classroom assessment techniques, albeit not from students' point of view (Cottell et al., 1998; Simpson Beck, 2011).

In the light of these findings, it is recommended to integrate classroom assessment techniques into the teaching of reliability and validity as a part of measurement and evaluation course. Also, it is thought important to examine the contribution of CATs to teaching of all topics of measurement and evaluation which particularly correspond to smaller literacy levels (Birgin 2007; Gürsoy, 2017; Karaman et al., 2014; Karaman, 2017) and unsatisfactory levels of competences (Çakan, 2004; Çalışkan et al., 2013; Demir Atalay, 2017; Evin Gencel et al., 2013; Gelbal et al., 2007; Kilmen et al., 2007; Pektaş, 2010; Sabancı et al., 2017; Yaralı, 2017; Yaman et al., 2011). Finally, bearing the limitations of the Turkish studies in mind, it would be again meaningful to discover the contributions of classroom assessment techniques to teaching in various disciplines and levels of learning.

References

Angelo, A. Thomas. & Patricia Cross. *classroom assessment techniques: a handbook for college teachers*. San Francisco: Jossey-Bass Publishers, 1993.

Baykul, Y. (2015). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Pegem Akademi.

Beard, V. (1993). Classroom assessment techniques (CATs): Tools for improving accounting education. *Journal of Accounting Education*, 11(2), 293-300.

Birgin, O., & Baki, A. (2009). An investigation of primary school teachers' proficiency perceptions about measurement and assessment methods: The case of Turkey. *Procedia-Social and Behavioral Sciences*, 1(1), 681-685.

Birgin, O., & Gürbüz, R. (2008). Sınıf öğretmen adaylarinin ölçme ve değerlendirme konusundaki bilgi düzeylerinin incelenmesi. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (20), 163-179.

Birgin, Osman. "*Sınıf öğretmeni adaylarının ölçme ve değerlendirme konusundaki okur-yazarlık düzeylerinin incelenmesi*," 16. Ulusal Eğitim Bilimleri Kongresi, 5-7 Eylül, Tokat 2007.

Büyüköztürk, Ş. (2001). *Deneysel desenler:* Öntest *sontest kontrol gruplu desen ve veri analizi*. Ankara: Pegem Akademi.

Byon, A. S. (2005). Classroom assessment tools and students' affective stances: KFL classroom settings. *Language and Education*, 19(3), 173-193.

Carduner, J. (2002). Using classroom assessment techniques to improve foreign language composition courses. *Foreign Language Annals*, 35(5), 543-553.

Cottell, P., & Harwood, E. (1998). Do classroom assessment techniques (CATs) improve student learning? *New Directions for Teaching and Learning*, 1998(75), 37-46.

Cross, T., & Palese, K. (2015). Increasing learning: Classroom assessment techniques in the online classroom. *American Journal of Distance Education*, 29(2), 98-108.

Çakan, M. (2004). Öğretmenlerin ölçme-değerlendirme uygulamaları ve yeterlik düzeyleri: İlk ve ortaöğretim. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 37(2), 99-114.

Çalışkan, H., Uymaz, M., & Tekin, D. (2013). Sosyal bilgiler öğretmen adaylarının ölçme ve değerlendirme yöntemlerine ilişkin yeterliklerinin değerlendirilmesi. *Adıyaman Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 6(14), 239-261.

Darling-Hammond, L., Wei, R. C., & Johnson, C. M. (2009). Teacher preparation and teacher learning: A changing policy landscape. *Handbook Of Education Policy Research*, 613-636.

Demir Atalay, T. (2017). Edebiyat öğretmen adaylarının ölçme ve değerlendirme genel yeterlik algısı. *International Journal of Language Academy*, 5(6), 389-400.

Duman, G., & Çetin, M. (2014). Çok kültürlü eğitim kavramına ilişkin öğretmen adaylarının *metaforik algıları*. 9. Uluslararası Balkan Eğitim ve Bilim Kongresi, 16-18 Ekim Trakya Üniversitesi, Edirne.

Eckert, C. M., D. J. Bower, K. S. Stiff, A. J. Hinkle, and A. R. Davis. 1997. Students' knowledge and faculty members' expectations: The case for classroom assessment. *Teaching Sociology*, 25 (2): 150-59.

Eminoğlu Küçüktepe, S. (2015). Analyzing the effect of instruction based on classroom assessment techniques in prospective teachers' Achievement. *Anthropologist*, 20(1,2), 223-231.

Field, A. (2009). *Discopering Statistics Using SPSS, Thrid Edition*. Sage Publication. Uk.

Gaeddert, B. K. (2003). Improving graduate theological instruction: Using classroom assessment techniques to connect teaching and learning. *Teaching Theology & Religion*, 6(1), 48-52.

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

Gencel, İ. E., & Özbaşı, D. (2013). Öğretmen adaylarının ölçme ve değerlendirme alanına yönelik yeterlik algılarının incelenmesi. *Elementary Education Online*, 12(1).

Giorgi, A., & Giorgi, B. (2003). *Phenomenology*. Sage Publications, Inc.

Goldstein, G. S. (2007). Using classroom assessment techniques in an introductory statistics class. *College Teaching*, 55(2), 77-82.

Gök, B., Erdogan, O., Özen-Altinkaynak, S., & Erdogan, T. (2012). Investigation of pre-service teachers' perceptions about concept of measurement and assessment through metaphor analysis. *Procedia-Social and Behavioral Sciences*, 46, 1997-2003.

Gronlund, N. E. (2002). *Assessment o f student achievement (7th ed.)*. Boston: Allyn & Bacon.

Gül, E. (2012). öğretmen *adaylarının ölçme ve değerlendirme okuryazarlığı*. XXI. Ulusal Eğitim Bilimleri Kongresi, Marmara Üniversitesi, İstanbul.

Günal, Y. (2014). *Rehber öğretmen adaylarının ölçme ve değerlendirme dersine ilişkin metaforik algıları*. 9. Uluslararası Balkan Eğitim ve Bilim Kongresi, 16-18 Ekim Trakya Üniversitesi, Edirne.

Gürsoy, G. (2017). Ölçme ve değerlendirme okuryazarlığı: Kavramsal bir analiz. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 6(1), 281-316.

Güven, S. (2001). *Sınıf öğretmenlerinin ölçme ve değerlendirmede kullandıkları yöntem ve tekniklerin belirlenmesi*. 10. Ulusal Eğitim bilimleri Kongresi, 7 – 09 Haziran, Bolu.

Güvendir, M. A., & Özkan, Y. Ö. (2016). Öğretmen adaylarının ölçme ve değerlendirme dersine yönelik algılarının metaforlar yoluyla belirlenmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (47), 91-105.

Howitt, D., & Cramer, D. (1997). *A guide to computing statistics with SPSS for Windows*. Prentice Hall.

Jawad, L. F., Majeed, B. H., & ALRikabi, H. T. S. (2021). The Impact of CATs on Mathematical Thinking and Logical Thinking Among Fourth-Class Scientific Students. *International Journal of Emerging Technologies in Learning (Online)*, 16(10), 194.

Karaman, P., & Şahin, Ç. (2014). Öğretmen adaylarının ölçme değerlendirme okuryazarlıklarının belirlenmesi. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 15(2), 175-189.

Karaman, P., & Şahin, Ç. (2017). Öğretmen adaylarının ölçme-değerlendirme okuryazarlıklarının mikro-öğretim yoluyla geliştirilmesi. *Electronic Turkish Studies*, 12(4).

Kilmen, S., Akın Kösterelioğlu, M. & Kösterelioğlu, İ. (2007). Öğretmen adaylarının ölçme değerlendirme araç ve yaklaşımlarına ilişkin yeterlik algıları. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 7(1), 117-127.

Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, 22(3), 255-276.

McNair, S. (2000). Holding Up a Mirror: Classroom Assessment Techniques in a Graduate Early Childhood Teacher Education Program. ERIC Document Reproductive Service No. ED447071.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook (Second Edi)*. Sage Publications.

Milli Eğitim Bakanlığı [MEB]. (2017). Öğretmenlik mesleği genel yeterlikleri. Ankara: Öğretmen Yetiştirme ve Geliştirme Genel Müdürlüğü.

Murdock, D. A. (2018). *The Effectiveness of a Classroom Assessment Technique (CAT) in the Calculus I Classroom* (*Doctoral dissertation)*. Texas A & M University.

Nartgün, Z. (2010). Sınıf içi durum belirleme tekniklerine dayalı öğretimin öğrencilerin araştırma teknikleri yeterlik düzeyleri üzerindeki etkisi. *Eğitim ve Bilim*, 35(157).

Nartgün, Z., & Uluman, M. (2009). Sınıf içi durum belirleme tekniklerine dayalı öğretimin öğretmen adaylarının başarısı üzerindeki etkisinin incelenmesi. *Uluslararası İnsan Bilimleri Dergisi*, 6(2), 626-650.

Pektaş, S. (2010). Öğretmen adaylarının ölçme ve değerlendirme yeterlik algılarının incelenmesi (yayınlanmamış yüksek lisans tezi). Abant İzzet Baysal Üniversitesi.

Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(1), 21-27.

Popham, W. J. (2002). Right task, wrong tool. *American School Board Journal*, 189(2), 18-22.

Saban, A. (2008). Okula ilişkin metaforlar. *Kuram ve Uygulamada Eğitim Yönetimi*, 55(55), 459-496.

Saban, A. (2009). Öğretmen adaylarinin öğrenci kavramina ilişkin sahip oldukları zihinsel imgeler. *Türk Eğitim Bilimleri Dergisi*, 7(2), 281-326.

Sabancı, O., & Yazıcı, K. (2017). Öğretmen adaylarının ölçme ve değerlendirmeye yönelik yeterlik algılarının incelenmesi. *Trakya Üniversitesi Eğitim Fakültesi Dergisi*, 7(1).

Schwarm, S., & VanDeGrift, T. (2002). *Using classroom assessment to detect students' misunderstanding and promote metacognitive thinking.* In Keeping Learning Complex: The Proceedings of the Fifth International Conference of the Learning Sciences (ICLS). Erlbaum, Mahway, NJ.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.

Simpson-Beck, V. (2011). Assessing classroom assessment techniques. *Active Learning in Higher Education*, 12(2), 125-132.

Soetaert, E. (1998). Quality in the Classroom: Classroom Assessment Techniques as TQM. *New Directions for Teaching and Learning*, 75, 47-55.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324-328.

Stronge, J. H., Tucker, P. D., & Hindman, J. L. (2004). *Handbook for qualities of effective teachers*. Ascd.

Taşgın, A., & Köse, E. (2015). Sınıf öğretmeni adaylarının hedef ve değerlendirme kavramlarına ilişkin metaforları. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30 (3), 116-130.

Tavşancıl, E., & Aslan, A. E. (2001). *Sözel, yazılı ve diğer materyaller için içerik analizi ve uygulama örnekleri*. Ankara: Epsilon.

Tunç, E. B., & Uluman, M. (2018). Öğretmen adaylarının açık uçlu ve çoktan seçmeli maddelere yönelik algılarının metaforlar aracılığıyla belirlenmesi. *Elektronik Sosyal Bilimler Dergisi*, 17(65), 180-194.

Walker, C. J. (1991). Classroom research in psychology: Assessment techniques to enhance teaching and learning. *New Directions for Teaching and Learning,* 46:67-78.

Welsh, M. E., & D'Agostino, J. (2009). Fostering consistency between standards-based grades and largescale assessment results. In T. R. Guskey (Ed.), *Practical solutions for Serious Problems in Standards-Based Grading*, 75–104. Thousand Oaks, CA: Corwin.

Xu, L., & Liu, G. (2009). The study of a method of regional environmental risk assessment. *Journal of Environmental Management*, 90(11), 3290-3296.

Yaman, S., & Karamustafaoğlu, S. (2011). Öğretmen adaylarının ölçme ve değerlendirme alanına yönelik yeterlik algı düzeylerinin incelenmesi. *Journal of Faculty of Educational Sciences*, 44(2).

Yaralı, D. (2017). Öğretmen adaylarının ölçme ve değerlendirmeye yönelik yeterlik algılarının incelenmesi (Kafkas Üniversitesi Örneği). *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 17(1).