ISTANBUL MEDENIYET UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF FOREIGN LANGUAGE EDUCATION
ENGLISH LANGUAGE TEACHING

Online Testing And Assessment In English As A Foreign Language Context: Teachers' Perspectives

Master's Thesis

İrem Gedil

July 2023

ISTANBUL MEDENIYET UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF FOREIGN LANGUAGE EDUCATION
ENGLISH LANGUAGE TEACHING

Online Testing And Assessment In English As A Foreign Language Context: Teachers' Perspectives

Master's Thesis

İrem Gedil

Supervisor

Prof. Dr. Selami Aydın

July 2023

TEZ JÜRİSİ ONAYI

İrem Gedil tarafından hazırlanan "Online Testing and Assessment in English as a Foreign Language Context: Teachers' Perspectives" başlıklı bu yüksek lisans tezi Yabancı Diller Eğitimi Ana Bilim Dalında hazırlanmış ve jürimiz tarafından kabul edilmiştir.

JÜRİ ÜYELERİ **İMZA**

Tez Danışmanı:

Prof. Dr. Selami Aydın

İstanbul Medeniyet Üniversitesi

Üyeler:

Dr. Öğr. Üyesi Ferdane Denkci Akkaş

İstanbul Medeniyet Üniversitesi

Dr. Öğr. Üyesi Tuba Arabacı Atlamaz

Abdullah Gül Üniversitesi

Tez Savunma Tarihi: 03.07.2023

BEYANLAR

Yazım ve Kaynak Gösterme Kılavuzu Beyanı

Danışmanlığımda yazılan bu tezin APA yazım ve kaynak gösterme kılavuzunda belirtilen kurallara uygun olarak yapılandırıldığı ve bu kılavuzun metin içi kaynak gösterme standartlarının bu tezde tutarlı olarak uygulandığı tarafımdan incelenerek teyit edilmiştir.

Prof. Dr. Selami Aydın

Etik İlkelere Sadakat Beyanı

Hazırladığım bu tezin tamamen kendi çalışmam olduğunu, akademik ve etik kuralları gözeterek çalıştığımı ve her alıntıya kaynak gösterdiğimi beyan ederim.

İrem Gedil

ÖZET

Yabanci Dil Olarak İngilizce Öğretiminde Çevrimiçi Ölçme Ve Değerlendirme: Öğretmen Perspektifi

Gedil, İrem

Yüksek Lisans Tezi, İngiliz Dili Eğitimi Anabilim Dalı

Danışman: Prof. Dr. Selami Aydın

Temmuz, 2023

Dil öğretiminde ölçme değerlendirme faaliyetleri, öğrencilerin bilgi ve becerilerini ölçmek; yeterliğini belirlemek, öğretimin başarılı olup olmadığını değerlendirmek, öğretim programının, kullanılan kitapların veya materyallerin etkinliklerini ölçmek veya bazı durumlarda öğretimin öğeleri konusunda mantıksal sonuçlar çıkarabilmek ve bunlara bağlı olarak kararlar alabilmek amacıyla yürütülmektedir (Aydın, 2004; Fulcher, 2013; Madsen, 1983). Geleneksel olarak etkin ölçme ve değerlendirme faaliyetlerinin belli başlı temel nitelikleri vardır ve her ölçme ve değerlendirme aracı bu nitelikleri belli ölçülerde sağlamalıdır. Bu temel nitelikler geçerlik, güvenirlik, kullanışlılık ve sınav güvenliğidir. Bu nitelikler bir ölçme ve değerlendirme aracının etkinliğini belirler ve dolaylı olarak da sınıf içi eğitim ve öğrenimi etkiler. Bu sebeple, çevrimiçi veya geleneksel farkı olmaksızın herhangi bir ölçme ve değerlendirme faaliyeti için bu niteliklerin belirlenmesi çok önemlidir.

Ölçme ve değerlendirme faaliyetleri, dil öğretiminde çok uzun zamandır kullanılıyor olmasına karşın değişen şartlar, teknikler ve amaçlar doğrultusunda değişime uğramış, dinamizm doğrultusunda gelişmiştir çünkü dil öğretimi, bir çok ilgili kişi barındıran, bu kişilerin, zamanın veya şartların farklılaşması doğrultusunda değişime uğrayan bir olgudur ve bu sebeple ölçme ve değerlendirme biçimleri de şartlar doğrultusunda değişecek,

gelişecek ve başkalaşacaktır (Bachman & Palmer, 2010). Teknolojik gelişmeler, bilgisayar kullanımının artması ve yaygınlaşması ve internetin ortaya çıkışı bu değişimi büyük ölçüde etkilemiş ve yön vermiştir. Bilgi ve iletişim teknolojilerinin zamanla gelişmesi ve yaygınlaşması, insanların çevrimiçi öğrenme metodlarını benimsemesini, ve bunun sonunda da çevrimiçi ölme ve değerlendirme yöntemlerinin ortaya çıkmasını mümkün kılmıştır (Çağlayan, 2021). Çevrimiçi ölçe ve değerlendirme yöntemlerinin zaman ve mekandan bağımsız oluşu (Yulianto & Mujtahid, 2021), esnek olması, kalabalık popülasyonlarla kullanmaya uygun olması, şekline bağlı olarak okunma kolaylığı, kaynak, zaman ve para israfını önlemesi ve bilgisayar kullanma konusunda çok rahat olan günümüz öğrencilerine çok uygun olması (Alghammas, 2020) gibi avantajlarını sıralamak mümkündür. Buna karşılık, sınav gözetmenliğinin zor veya etkisiz olması, artan kopya çekme olasılığı, güvenlik sorunları, yavaş internet bağlantısı, internet yoksunluğu, öğrencilerin veya öğretmenlerin teknoloji kullanma konusunda yaşadıkları zorluklar gibi sebepler çevrimiçi ölçme ve değerlendirme yöntemlerini kulllanmayı zorlu hale getirmektedir (Astiandani & Anam, 2021).

Çevrimiçi ölçme ve değerlendirme, görece yeni bir kavram olduğundan, bu konu üzerinde alanyazın çalışmaları oldukça sınırlı, ve dar bağlamlıdır (Butler-Henderson & Crawford, 2020). Günümüzde bunların çoğu, sonuçlarından direkt olarak etkilendikleri için en temel paydaşlardan olan öğrencilerin görüşlerini incelemektedir (Biggs & Tang, 2007); bir kısmı yine önemli etkileyen ve etkilenenlerden oldukları için hizmet öncesi öğretmen algılarına, bir kısmı da geleneksel ölçme değerlendirme yöntemlerinin çevrimiçi ölçme ve değerlendirme yöntemleri ile kıyasına bakmaktadır. Ne var ki sistemin temel kullanıcıları ve uygulayıcıları olmalarına karşın, yabancı dil öğretimi bağlamında Türkiye'de görev yapan İngilizce öğretmenlerinin çevrimiçi ölçme ve değerlendirme konusundaki görüşlerine yer veren araştırma sayısının azlığı dikkat çekicidir. Sistemin

temel kullanıcısı olarak öğretmenlerin görüşleri büyük önem arz etmektedir çünkü onların görüş ve algıları, tutumlarını belirleyecek ve tutumları ders ortamına yansıyacaktır. Bu sebeple, bu derece büyük etkisi olan bir faktörün derin biçimde analiz edilmesi, etkileyen faktörlerin bulunması gerekmektedir.

Alanyazın incelendiğinde, mevcut bulunan araştırma sonuçlarına göre, öğretmenlerin çevrimiçi ölçme ve değerlendirme konusunda çeşitli görüşlere sahip olduğu görülmektedir. Birçok araştırmada, öğretmenlerin çevrimiçi ölçme ve değerlendirme konusunda olumlu algı ve tutumlara sahip olduğu görülmüştür (Baleni, 2015; Chien et al., 2014; Fageeh, 2015; Fitriyah & Jannah, 2021; Küppers & Schroeder, 2020). Bu olumlu tutumun sebebinin çevrimiçi değerlendirmenin sınav kaygısı ve gerginliğini azaltması (Baleni, 2015; Öz, 2014b;), esnek ve çevre dostu olması (Asma, 2021), ve eğitim öğretim faaliyetlerine olan olumlu etkileri (Fageeh, 2015) olduğu belirtilmiştir. Fakat yine birçok çalışmada, öğretmenlerin çevrimiçi ölçme ve değerlendirme konusundaki en büyük endişesinin sınav güvenliği olduğu görülmüştür (Alghammas, 2020; Fitriyah & Jannah, 2021; Küppers & Schroeder, 2020; Meccawy et al., 2021; Rogers, 2006; Sa'di et al. 2021). Bu çalışmalarda öğretmenlerin çevrimiçi ölçme ve değerlendirme faaliyetleri sırasında öğrencilerin daha çok ve sık kopya çektikleri, sistemin buna daha açık olduğu ve bu kopya çekme faaliyetlerinin bir çok tür ve biçimde olduğunu düşündükleri ortaya konmuştur. Bunun yanısıra, teknik aksaklıklar (Alghammas, 2020), öğretmenlerin teknolojik bilgi ve deneyim düzeyinin düşük olması (Joshi, 2020; Nowroozi, 2021), bilgisayar veya ekran yorgunluğu (Asma, 2021), çevrimiçi ölçme ve değerlendirme sistemine güvensizlik (Rollim & Isaias, 2018), ve geleneksel ölçme ve değerlendirme biçimlerini tercih etme gibi başka olumsuz noktalar da belirtilmiştir (Yulianto & Mujtahid, 2021). Aynı zamanda kimi çalışmalarda, öğretmenlerin çevrimiçi ölçme ve değerlendirme ile ilgili tutumlarının cinsiyete göre değişip değişmediği de incelenmiş, ve farklı sonuçlar bulunmuştur.

Sonuç olarak, çevrimiçi ölçme ve değerlendirme konusunda öğretmen algılarının belirlenmesi, halihazırda bulunan çalışmaların kısıtlılığı sebebiyle görece yeni olan bu alan için büyük önem arz etmektedir. Bu çalışma, Türkiye'de üniversite düzeyinde çalışan öğretmenlerin çevrimiçi ölçme ve değerlendirme yöntemleri ile ilgili genel tutum ve yaklaşımlarını, duyuşsal faktörler, geçerlik, güvenirlik, güvenlik, uygulanabilirlik ve eğitim ve öğrenim üzerine etkileri bağlamlarında incelemeyi amaçlamıştır. Çalışmanın diğer amacı da, öğretmenlerin çevrimiçi ölçme ve değerlendirme algı ve yaklaşımlarının yaş, cinsiyet, öğretmenlik deneyimi, mezun olunan bölüm, eğitim düzeyi, çalışılan kurum, yönetimsel veya işlevsel bir görevi bulunup bulunmaması, öğretmenlerin kendi beyanlarına göre bilgisayar becerileri gibi demografik değişkenlere göre değişip değişmediğinin belirlenmesidir. Tanımlayıcı nitelikteki bu çalışma için, Türkiye'de çeşitli üniversitelerin İngilizce hazırlık programında çalışan 302 İngilizce öğretmeninden çevrimiçi bir anket yardımıyla veri toplanmıştır.

Uygulanan anket, Dermo (2009) tarafından geliştirilmiş ve ilk kullanımında öğrencilerin çevrimiçi ölçme ve değerlendirme algı ve tutumlarını belirlemek amacıyla kullanılmıştır. Sonrasında, aynı anket çeşitli ülkelerde ve bağlamlarda öğrenci ve öğretmenlerin çevrimiçi ölçme ve değerlendirme tutum ve algılarını belirlemek için kullanılmış, çeşitli sonuçlar elde edilmiştir. Bu araştırma bağlamında aynı anket, orijinalinden öğretmenlere uygun olacak şekilde değişiklik yapılarak kullanılmıştır ve iki bölümden oluşmaktadır. Birinci bölüm, öğretmenlerin demografik bilgilerini içeren sorulardan ve ikinci bölüm de öğretmenlerin genel çevrimiçi algılarını ve duyuşsal faktörler, geçerlik, güvenirlik, güvenlik, uygulanabilirlik ve çevrimiçi ölçme ve değerlendirmenin eğitim ve öğrenim üzerine etkileri ile ilgili algı ve fikirlerini ölçmeyi amaçlayan 30 Likert tipi sorudan oluşmaktadır. Anket katılımcılarına anket kişisel e-postalar yoluyla iletilmiş ve bir ay kadar bir süre içinde veriler toplanmıştır. Toplanan veriler SPSS veri inceleme

programı ile incelenmiş ve verilere göre anlamlı sonuçlar elde edilmiştir. Öncelikle, ankete katılan katılımcıların büyük bir çoğunluğunun çevrimiçi ölçme ve değerlendirme konusunda nötr,yani ne olumlu ne de olumsuz bir bakış açısına sahip olduğu saptanmıştır. Benzer biçimde, öğretmenlerin algı ve tutumları çevrimiçi ölçme ve değerlendirmenin alt bağlamları olan duyuşsal faktörler, geçerlik, güvenirlik, kullanışlılık, ve güvenlik alanlarında da nötr olarak bulunmuştur. Yalnızca çevrimiçi ölçme ve değerlendirmenin eğitim ve öğrenim üzerine etkileri alanında öğretmenlerin olumu görüş bildirdiği görülmüştür. Bu çalışma, daha önce belirtilen öğretmenlerin pozitif tutum ve yaklaşım sergilediği çalışmalardan bu yönüyle ayrılmaktadır. Bunun yanısıra, sınav güvenliği de en fazla olumsuz görüş bildirilen alan olmuştur ve bu bulgu ile daha önce bahsedilen diğer bir çok araştırma bulgusunun uyumlu olduğu anlaşılmıştır. Bir diğer bulgu da çeşitli demografik etkenlerin öğretmenlerin çevrimiçi ölçme ve değerlendirme algı ve tutumları üzerinde etkili olmasıdır. Araştrma sonuçlarına göre, cinsiyet, yaş, öğretmenlik deneyimi ve öğretmenlerin kendi beyan ettikleri bilgisayar becerileri faktörlerinin onların çevrimiçi ölçme ve değerlendirme algı ve tutumları üzerinde etkili olduğu; çalışılan kurum, mezun olunan bölüm, eğitim düzeyi ve kurum içerisinde yönetimsel veya işlevsel görevler alıp almadığının bu algı ve fikirleri etkilemediği görülmüştür.

Araştırma sonuçları öğretmenlerin çevrimiçi ölçme ve değerendirme algı ve görüşlerinin cinsiyetlerine göre değişebildiğini göstermiştir. Buna göre, erkek öğretmenlerin, kadın öğretmenlere göre çevrimiçi ölçme ve değerlendirmeye daha olumlu algı ve tutum geliştirdiği, bir çok anket maddesinde de kadınlardan daha olumlu görüş bildirdiği gözlenmiştir. Bir diğer etkili faktör olan yaş değişkenine göre de, 31-40 yaş aralığında olan öğretmenlerin çevrimiçi ölçme ve değerlendirme noktasında 50 yaş ve üstü öğretmenlere göre daha olumlu algı ve tutumlarının olduğu, bu sunucun neredeyse diğer tüm alt bağlamlarda da görüldüğü; ve hatta 50 yaş ve üstü

öğretmenlerin 20-30, 31-40 ve 41-50 yaş gruplarına göre bir çok anket maddesinde daha olumsuz bakış açısında olduğu izlenmiştir. Bu sonuçlara göre, 50 yaş altı öğretmenelerin genel olarak 50 yaş ve üstü öğretmenlere göre çevrimiçi ölçme ve değerlendirme hususunda daha olumlu bir bakış açısına sahip oldukları söylenebilir. Bir diğer etkili değişken olan öğretmenlik deneyimi de zaten yaş ile demografik olarak doğru orantılı olan bir değişken olduğundan benzer sonuçlar ortaya çıkarmış; on yıl ve altı deneyim sahibi olan öğretmenler hem genel algı ve tutumlarında, hem de neredeyse tüm alt bağlam algı ve tutumlarında 30 yıl ve üstü deneyimi olan öğretmenlere göre daha olumlu bakış açısı geliştirmiş; bu farkı bir çok anket maddesinde de yansıtmışlardır. Buna göre, 30 yıldan daha az deneyimli öğretmenlerin, 30 yıldan fazla deneyimli öğretmenlere göre çevrimiçi ölçme ve değerlendirme konusuna daha olumlu baktıkları bir gerçektir. Son önemli farklılıklar gösteren değişken olan bilgisayar becerileri öğretmenlerin kendi kendilerini değerlendirdikleri bir soru ile ortaya konulmuş olup öğretmenlerin beyanına dayanmaktadır. Bu değişkenin analizi, bilgisayar becerilerini mükemmel olarak niteleyen öğretmenlerin bilgisayar becerilerini yeterli olarak niteleyen öğretmenlere göre çevrimiçi ölçme ve değerlendirme konusunda çok daha olumlu bakış açısına sahip oldukları, bu sonucun neredeyse tüm alt bağlamlara yansıdığı ve hatta bir çok anket maddesinin incelenmesinde de ortaya çıktığı görülmüştür. Bu, öğretmenlerin bilgisayar becerileri geliştikçe, çevrimiçi ölçme ve değerlendirme konusunda daha olumlu bir bakış açısına sahip olacağını düşündürmektedir.

Tüm bunların ışığında, çevrimiçi ölçme ve değerlendirmenin gelişen ve değişen teknolojiiyle birlikte daha da yaygınlaşacağı varsayılırsa, öğretmen algı ve tutumlarını geliştirmek için araştırmacılara, öğretmenlere ve kurumlara belli görevler düşmektedir. Öncelikle araştırmacılar öğretmenlerin karar alma mekanizması içindeki önemli yerini kavramalı, ve çevrimiçi ölçme ve değerlendirme konusunda öğretmen algı ve tutumlarını farklı bağlamlarda

ve farklı öğretmen popülasyonlarıyla araştırmaya zaman ayırmalıdır. Bu araştırmaları nicel araştırmalarla da desteklemeli, öğretmen algı ve tutumları konusunda mümkün olduğunca net bakış açıları elde etmeyi hedeflemelidir. İkinci olarak, öğretmenler çevrimiçi ölçme ve değerlendirme konusunda bilgi sahibi olmalı ve avantaj ve dezavantajlarını iyi bilmelidirler. Bu konularda yeterince bilgi sahibi olurlarsa, daha doğru bir bakış açısı elde edebilirler. Bunun yanı sıra, öğretmenlerin bilgisayar becerisi geliştikçe, çevrimiçi ölçme ve değerlendirme konusundaki algı ve tutumları da gelişmekte ve endişe seviyesi azalmaktadır. Bu sebeple, öğretmenlerin bilgisayar becerilerini gerekli meslekiçi eğitimlerle geliştirmesi gerekmektedir. Son olarak kurumlar da öğretmenlerin bu ihtiyaçlarını göz önünde bulundurmalı, onlara çeşitli eğitimler vererek çevrimiçi ölçme ve değerlendirme konusunda bilgi sahibi olmalarını sağlamalı, ve gerekli kurslarla onların bilgisayar becerilerini desteklemelidirler. Ayrıca, öğretmenlere sınav öncesinde, sınav sırasında ve sonrasında gerekli teknik desteği sağlamalı ve son teknolojik yatırımlarla sınav sistemlerini geliştirmelidirler. Bunun yanı sıra, araştırma sonuçlarına göre öğretmenlerin en çok endişe ettiği ve olumsuz bulduğu nokta olan sınav güvenliği de alınacak bir takım önlemlerle iyileştirilebilir. Ayrıca gelişmiş sistemlerle sınav güvenliği sağlanırsa, öğretmenlerin sisteme olan güveni de artacaktır.

Bu çalışmanın çeşitli kısıtlamaları da mevcuttur. Öncelikle, çalışma belli bir öğretmen grubu ile yapılmıştır ve bu sebeple tüm öğretmen gruplarına genelleme yapılması doğru olmayabilir. Benzer biçimde, araştırma evreni Türkiye'de çalışan öğretim görevlilerinden oluşmaktadır, bu sebeple, araştırmanın diğer ülkelerdeki külürel farklar da göz önüne alınarak yorumlanması gerekmektedir. Bununla birlikte, araştırma verisi yalnızca nicel yöntemlerle toplanmıştır. Bu verilerin nitel yöntemlerle de desteklenmesi, veri analizi ve yorumlaması için çok daha sağlıklı bir yöntem olacaktır. Ayrıca bu araştırma algı ve tutumların çalışıldığı bir araştırmadır ve algı ve tutumlar

kişisel görüşler olduğundan, nicel bir çalışma ile ölçülmesi zor olabilmektedir. Aynı zamanda, algı ve tutumların çeşitli kaygılarla tam olarak doğru şekilde yansıtıldığından emin olmak da aynı şekilde zordur. Bu bağlamda, araştırmayı dikkatle yorumlamak gerekecektir.

Anahtar Kelimeler: EFL, çevrimiçi ölçme ve değerlendirme, öğretmen algı ve tutumları, e-değerlendirme, dijital ölçme ve değerlendirme.

ABSTRACT

Online Testing and Assessment In English As A Foreign Language Context: Teachers' Perspectives

Gedil, İrem

Master's Thesis, English Language Teaching

Supervisor: Prof. Dr. Selami Aydın

July 2023

Being one of the indispensable elements of language teaching, testing and assessment play a vital role in identifying whether teaching has been effective and the aims of the program has been reached (Madsen, 1983). To this end, institutions assess the students through various methods at regular time intervals. As information and communication technologies have been advancing and becoming more widespread, it has enabled people to adopt online learning and this has given rise to online assessment (Çağlayan, 2021). As online assessment is a rather new phenomenon, available research regarding it is limited (Butler-Henderson & Crawford, 2020), and most of the research in literature today examines the views of students regarding online assessment as they are one of the most important stakeholders of exams. On the other hand, although they are the basic users and practitioners of the system, studies on the views and perspectives of instructors on online assessment in English as a foreign language context are limited in number. With these facts in mind, this study aims to explore English instructors' perspectives of online assessment in tertiary educational institutions in Turkey with regard to their general views on online assessment, along with their views on affective factors, validity, reliability, security, practicality and the impacts of online assessment on teaching and learning. In this descriptive study, the data was collected from 302 English instructors working at

English preparatory schools in various universities in Turkey through an online survey. The survey collected demographic information of the participants and gathered their views on abovementioned aspects with 30 Likert scale survey questions. The data collected were analyzed through SPSS statistical software. The analysis of the data revealed two important results: First, the overall perspectives of instructors on online assessment were neutral. Second, the results indicated that instructors' perspectives of online assessment differed significantly according to their gender, age, teaching experience years, and their self-reported computer expertise. On the other hand, their educational background, department of graduation, the institution they worked at and holding administrative or office duties did not change instructors' perspectives significantly.

Keywords: EFL, online assessment, teacher perspectives, e-assessment, digital assessment.

ACKNOWLEDGEMENTS

This thesis has been the product of endless support and encouragement from many people in my life. Therefore, I would like to express my gratitude to them all.

First and foremost, I would like to express my heartfelt regards to my supervisor Prof. Selami Aydın for his invaluable support and guidance, endless patience, and constant positive encouragement. Without his never-ending help and support, this thesis would not have been possible.

I would also like to express my deep gratitude to Assoc. Prof. Özge Cengiz and Asst. Prof. Ferdane Denkci Akkaş for their valuable guidance and help throughout my master's education. I would like to extend my gratitude to Dr. Tuba Arabacı Atlamaz for her useful suggestions and encouraging attitude.

Thanks should also go to my kind friends who supported me through the process of researching and writing this thesis. I would like to express my special thanks to my dearest friend Merve Yıldırım, whose continuous support, unconditioned love, and encouragement were invaluable.

I would also like to extend my thanks to the Scientific and Technological Research Council of Turkey (TUBITAK) for providing financial support within the scope of the 2210-A National Scholarship Program for Master's Students.

I also would like to thank my dearest father Mustafa Abi, my sister Çiğdem, and all my family members who helped me walk my way. Last but not least, I owe special thanks to my dear husband Abdullah, whose endless support, help, encouragement, unconditioned love, and understanding greatly helped me on my way, and my dear children Zeynep and Oğuz, who supported me in all ways beyond expectations from their young age.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EFL: English as a foreign language

IT: Information Technology

IRT: Item Response Theory

SPEAQ: Student Perceptions of e-Assessment Questionnaire

SPSS: Statistical Package for Social Sciences

TAM: Technology Acceptance Model

1. INTRODUCTION

The introduction section consists of fice subsections. In the first subsection, background to the study is explained with regard to importance of testing and assessment, online testing and assessment, teacher perceptions and the importance of teacher perceptions in online assessment. Then, the problem is discussed in detail. After that, the purpose of the study is explained together with the research questions. Then, the significance of the study for different parties is given and finally limitations of the study is explained in detail.

1.1. Background To The Study

Testing and assessment play a crucial role within the English as a Foreign Language (EFL) context, exerting indispensable significance for both students and educators through varied means. First, assessment is essential for learners as it increases their motivation and interest, and eventually helps students learn a language (Madsen, 1983). When teachers employ meaningful and trustworthy assessment methods, students will be more willing to learn the language and enjoy a sense of achievement, thereby facilitating their overall learning process. In addition, when students are aware that they will be evaluated, they will study the language more willingly, thus, yielding a positive impact on their overall learning outcomes. Second, testing and assessment are very important for teachers to collect information on the language capabilities of learners (Hughes, 1989). With the help of tests and evaluation methods, teachers will be able to identify students' level of language proficiency, their strengths, and weaknesses and will be able to measure if learners have achieved their goals. Therefore, they will be able to get to know the students and their needs easily. Finally, according to testing and evaluation results, teachers and administrators can make educational decisions in the right direction (Hughes, 1989). This way, the teaching program, books and materials in use, course content, and the teaching method will be effectively evaluated and re-adjusted if necessary (Aydın, 2004). For these reasons, testing and assessment have been the backbone of language teaching for enlightening the teaching and learning process and providing the opportunity to improve them. Without assessment, it will not be possible to evaluate learner responses to educational activities (Pehlivan Şişman & Büyükkarcı, 2019).

Online testing and assessment in the EFL context are also significant for three fundamental reasons. First, online testing and assessment fit well into the 21st-century language teaching context since it is the natural outcome of changing and evolving education systems with improving technology. Today's students, called Generation Z learners or I-Gens (Rothman, 2016), prefer the involvement of technology in their language learning experience and as online assessment is considered a more accessible and suitable system for today's students (Hichour, 2022; Prensky, 2010), this preference should be reflected in language assessment activities (Appiah & van Tonder, 2018). Second, online testing and assessment provide increased practicality, logistic efficiency and reliability in language assessment (Long et al., 2018). It allows the test taker to take the test in any location and at any time, increasing the flexibility for taking the test. Moreover, when marking is done by automatic scoring, online assessment increases practicality by saving time and effort and accuracy by utilizing computer programs for marking. It also provides immediate reporting of the results to stakeholders, thereby saving time. In addition, since test developers can upload and update test items easily, with little time and location restrictions, it eases the test developers' work (Long et al., 2018). The final reason why online testing and assessment is essential is that it provides the opportunity to continue testing and evaluation activities in the EFL contexts even in times of crisis (Alghammas, 2020; Hichour, 2022). With the help of online assessment methods, language learning and assessment activities can continue uninterruptedly even when face-to-face learning and assessment have to stop.

How language teachers perceive testing and assessment is another important aspect for three main reasons. The first reason is related to the role of teachers as decision-makers. As one of the most important relevant parties of testing and assessment, language teachers use the information they gain from the tests on students' performances in the decision-making process. They see if the instruction has been effective, language learning objectives have been met, language lessons are at the right level, or if they need to focus more on certain materials or not with the help of tests and assessment tools (Madsen, 1983). Thus, teachers may evaluate and reshape teaching and learning activities according to the test results of learners (Bachman & Palmer, 1996). Second, EFL teachers' role in interpreting assessment results is vital as they might end up "teaching to the test", which

is generally defined as the backwash effect of the test (Shim, 2009). The backwash of the assessment may directly or indirectly impact the teaching methods teachers employ in the EFL contexts, meaning that teachers' attitudes and perceptions about language testing and assessment may influence their work in class. Another reason for the importance of teachers' perceptions of assessment is their dual role in teaching and assessing (Shohamy et al., 2008). How language teachers perceive assessment might affect how they assess students (Black & Wiliam, 2010; Scarino, 2013). Thus, the responsibility of teachers in the assessment context is highly important in order not to misjudge or poorly grade students. When teachers have faith in the assessment method, they assess students more fairly (Xerri & Vella Briffa, 2018). One final reason for the importance of their perceptions in testing and assessment is their ability to shape learner behavior. Since teachers need to teach learners how to develop self and peer-assessment strategies for better learning, they must understand the assessment well. Moreover, teachers have the power to shape how learners feel about an assessment as learners might develop test anxiety which might prevent them from showing their true performance or feel comfortable about the assessment which might affect their learning positively according to the comments of their teacher (Brown, 2003).

Teachers' perceptions of online testing assessment in the EFL context are also very important for three reasons. First, EFL teachers need to adapt to technological improvements since they are teaching and assessing today's tech-savvy students who automatically need the involvement of technology to be interested (Mahbub, 2020). As language teachers need to grab students' attention, they need to make use of online assessments. However, if they do not believe in the effectiveness or usefulness of the system, they cannot appeal to students. Therefore, knowing teachers' perceptions regarding online testing and assessment in the EFL context is paramount. Second, to improve assessment, it is necessary to identify the needs and requirements of teachers with regard to online assessment methods (Gamage et al., 2020). Improving assessment or catering to their needs may not be possible without asking them about their perceptions, wants, and needs. Third and last, it is important to discover teachers' perceptions of online assessment in the EFL context to see how well their opinions match with the principles of language learning and teaching in teachers' minds. Since their perceptions greatly affect

their performance in class, it might be important to understand what teachers think of online assessment and make changes in the curriculum, assessment methods, teaching methods, and the teaching program accordingly (Balaman & Tiryaki, 2021).

1.2. Statement of the Problem

Testing and assessment is a complex area with many variables and principles, and it is of utmost importance for teachers to understand these variables and follow these principles to reach healthy conclusions for all stakeholders involved in the process (Yavuz Kırık, 2008). However, due to its complexity, many tensions and issues arise in this area. The first tension is about the validity of the assessment which relates to the link between what is being taught and what is assessed. As communicative language teaching suggests, all skills of the language should be tested, whereas objective testing may not bring in any language production and thus decrease the validity of the assessment (Morrow, 1981). When subjective tests and assessments are taken into account to cater to this need, the second problem, reliability concerns arise. When there is a discrepancy between the correct score and the observed score of a test because of many different variables such as factors related to the test taker, the test administration, the scorer, or the test itself, there can be a reliability problem (Farhady, 2021). Certainly, a test with unreliable scores will lead to undesirable consequences in the teaching and learning context. A third tension is about the washback of the test, and the effect of the test and assessment on teaching and learning, which might be beneficial or harmful. The harmful effects of the assessment on instruction might appear as limiting instruction time, changing the materials and curriculum according to the test, limiting teacher freedom, or changing teaching methods according to the test material (Göktürk Sağlam, n.d., Chapter 1, p. 15). Another tension is the security of the test and test procedures. It is of utmost importance for institutions to ensure test security as when the stakes are high for a test and cheating occurs, decisions to be made according to test results can be incorrect (Fulcher & Davidson, 2007).

As online testing and assessment in the EFL context have recently gained popularity, its problems are gradually emerging. One of the most commonly referred problems related to online testing and assessment is the issue of security (Arnold, 2016; Mellar et al., 2018; Rogers, 2006). It is believed that language students find many interesting ways to cheat or commit plagiarism during online testing and assessment

practices, negatively affecting the reliability of the assessment (Rogers, 2006). Additionally, many teachers believe that online tests and assessments make both cheating and plagiarism easier (Mellar et al., 2018). The second problem of online assessment is the issue of validity. As many online tests include objectively marked items such as multiple choice items, true/false items, or fill-in-the-blanks types of items, opportunities to foster students' critical thinking abilities are overlooked. Thus, many of the online EFL assessment tasks and tests are found to be disempowering since students who do not add any ideas, improve existing opinions, or devise new ways of thinking, become passive participants in activities (Öz, 2014b). Finally, the issue of accessibility and practicality is also a problem for online testing and assessment. Lack of necessary equipment or internet connection can lead to serious problems during the process of testing and assessment tasks (Alruwais et al., 2018). Sometimes, students tend to feel anxious because of the Internet connection, Internet speed, or software problems such as unexpected updates, system failures, or overloaded systems (Khan & Khan, 2019). In addition, complicated test procedures which might require IT support might result in serious consequences for the stakeholders (Fitriyah & Jannah, 2021).

An important issue about testing and assessment is that despite their significant role as item writers, assessors, or decision-makers, teachers' perception of testing and assessment in the EFL context is usually ignored (Sevilen, 2021). Unfortunately, when their opinions about testing and assessment practices are not given enough importance to, testing and assessment activities might not lead to better learning outcomes in the language learning context, contrary to what is expected of them. Since EFL teachers are to choose learning materials, make up the curriculum, or define learning objectives, their perceptions and conceptions play a very important role in making those decisions, and as the research on the issue is limited in number (Mede & Atay, 2017), the EFL learning context might be deeply affected by this lack of insight negatively. Secondly, the understanding which suggests that not all teachers participate in item writing activities and they do not need to have an understanding of assessment-related issues is a problematic approach since teachers need to be highly aware of assessment issues as one of the basic stakeholders of testing and assessment, even if they are not test writers (Sevilen, 2021). That is because teachers' understanding of assessment activities

fundamentally affects classroom learning and teaching activities (Sahinkarakas, 2012). Thus, raising awareness on the issue of teacher perceptions regarding assessment in the EFL context is very important.

As said before, since online testing and assessment in the EFL context is a new practice in many institutions, it has brought about many challenges and uncertainties (Gamage et al., 2020), most of which are directly related to teacher roles and responsibilities. As the process involves many uncertainties, it causes many differences in teachers' ideas. Thus, the problem of now knowing how EFL teachers feel about these uncertainties is an issue in terms of the effectiveness, efficiency and safety of online testing and assessment procedures. However, the general feelings and attitudes of EFL teachers are not known about the issue because of lack of research (Rea-Dickins, 2004). When the teachers' overall perceptions are not known, how much they accept this new phenomenon is also subject to doubt. As the acceptance level of new technology is unknown, it is impossible to understand the general attitude toward the new procedures (Al-alak & Alnawas, 2011). This can also be valid for the components of validity, reliability and the effects of assessment on learning and teaching. In addition, EFL teachers' perception of online testing and assessment is highly affected by their computer expertise (Alruwais et al., 2018), and existing studies on the issue reflect conflicting results (Öz, 2014a). For this reason, it is crucial to explore how teachers perceive online testing and assessment concerning their computer expertise.

1.3. Purpose of the Study

In its general sense, language teaching assessment aims to gather information about different aspects of educational processes to make meaningful decisions about plausible action plans to improve teaching and learning (Carol, 1961 as cited in Fulcher, 2010). To this end, as one of the important stakeholders of the assessment process, teachers' role in it should not be underestimated but rather valued and investigated. Their perception and understanding of assessment will play a significant role in the decision-making process as when they develop a solid understanding and ownership of the assessment, they will affect the whole process by making beneficial decisions for learners and also by improving learner and public acknowledgment through raising awareness (Xerri & Vella Briffa, 2018). Moreover, existing literature shows that how teachers

conceive assessment deeply affects how they behave in the classroom, directly impacting learning and teaching (Sahinkarakas, 2012). As the core of the language learning process, classroom activities might be affected by teachers' perceptions, and their opinions on assessment should be taken into consideration. In short, when institutions apply online testing and assessment procedures in language classes, it is vital to identify teacher perceptions of the new assessment procedures (Alghammas, 2020).

With these concerns in mind, this study aims to discover teachers' perceptions of online assessment practices in the EFL context. This study hopes not only to raise awareness of EFL teachers' perceptions but also to help decision-makers other than the teachers in tertiary education contexts such as test writers and school administrators, and lead the way toward the further analysis of online assessment-related issues in the EFL context. As online assessment is rather a new phenomenon in the Turkish EFL context, this study also aims to identify how teacher perceptions toward online assessment differ in the Turkish university-level EFL context according to gender, age, teaching experience, graduation department, highest completed educational degree, type of institution, institutional role and computer literacy variations and means to arouse further online assessment research.

1.3.1. Research Questions

As teachers' perceptions of online testing and assessment are crucial for many reasons discussed above, it is important to explore their overall perceptions toward online testing and assessment and their perceptions of online testing and assessment in terms of basic testing and assessment principles. It is also meaningful to be able to discover if their points of view change according to their gender, age, teaching experience, graduation degree, graduation department, type of their institution, their role in their institution, and their self-perceived proficiency in computer use. Therefore, the research questions of this study are:

1. How do EFL instructors perceive using online testing and assessment?

2. Is there a significant difference between the perceptions of EFL instructors toward online assessment with regard to certain variables?

1.4. Significance of the Study

Testing and assessment are basic and fundamental aspects of language teaching. As the executives of testing and assessment prartices and procedures, teachers' perspectives of testing and assessment play a significant role in different aspects. As a relatively new application, online testing and assessment is very important in today's technology-integrated world, and understanding teachers' perceptions on online testing and assessment is necessary for several reasons stated above. Yet, there are many issues both in traditional testing and assessment and in online assessment, and in order to enlighten these issues, a solid and sound analysis of teachers' perceptions of online assessment is vital.

Thus, this study will contribute to the existing literature by providing scientific data on a rather immature field of research, which is online testing and assessment, collected from a specific group of teachers who work in the field of EFL. This study may also help relevant parties such as teachers, administrators, curriculum developers, and item writers working in the EFL context to make educated decisions regarding online testing and assessment issues. The results of the data and the conclusions based on it were analyzed meticulously and, therefore, may be helpful in the future with regard to planning and administration of online tests and assessments in the EFL context in a more relevant manner. Finally, by shedding light on how teachers perceive online assessment in the EFL context, this study may open the gate for further research on online assessment practices in the EFL context.

1.5. Limitations

There are several limitations of this study. First of all, the data of this study are collected from 302 participants who worked in a specific context in Turkey. This is a limited number compared to the whole teacher population working at university level EFL teaching contexts. Thus, the results cannot be generalized to all teachers. Moreover, the data are collected in Turkey and therefore, caution needs to be taken while commenting on the results of the study. Cultural differences of the Turkish context with other countries toward the use of technology should be considered while analyzing the results of this study. Secondly, only quantitative data is used to come up with descriptive results. However, quantitative data collection methods such as surveys and questionairres should

be supported by interviews or observations in order to get more reliable results (Seliger & Shohamy, 1989). Thus, the current study lacks qualitative data support in this sense. This could not be done in this study because of time constraints. Another limitation is that the data is collected in a limited time and thus, does not represent the participants' ideas for throughout their lives. They might gain more insight into the issue, improve their computer literacy and change their opinions in the future. A fourth limitation to this study is the challenge of studying the perspectives of instructors as perspectives, beliefs, perceptions and attitudes are personal values which are hard to measure and explain, especially with a quantitative study. Moreover, instructors may have restrained from giving their honest opinions about online assessment for various personal reasons. Therefore, the data should be approached with caution in this respect.

2. RELATED LITERATURE

This section includes two subsections: the theoretical framework and the literature review. The first subsection delineates language testing and assessment, test types, basic testing and assessment principles, theories of language testing, and then focuses on online assessment, theories behind it and advantages and constraints of online testing and assessment. The second subsection goes over the relevant studies on online assessment in language teaching with regard to teacher perspectives, student perspectives and studies analyzing both teacher and student perspectives in the same context worldwide and in Turkey.

2.1. Theoretical framework

This section explains the commonly used testing terminology and the differences between them with reference to test types and assessment methods. Then, basic test principles are explained in detail. Afterward, theories of testing is explained. The following section focuses on online testing and assessment in definition, theories explaining online testing and assessment and discusses advantages and disadvantages of online testing and assessment.

2.1.1. Basic Concepts in Language Assessment

Before explaining the main features of good language tests, it is necessary to clarify the basic terminology in assessment. The concepts of assessment, measurement, evaluation and testing may be used interchangeably in literature; however, they possess different meanings. Therefore, in this section, these commonly used but confused concepts, assessment, evaluation, measurement, and testing are explained in detail.

2.1.1.1. Assessment, Measurement, and Evaluation

According to many (Bachman, 2004; Brown, 2003; Green, 2014; Linn & Miller, 2005), assessment, as a comprehensive concept, involves a cyclical process encompassing testing and measurement within. The main aim of assessment is to systematically collect information on an individual's abilities, skills or knowledge and make judgments or interpretations on the individual’s use of language (Bachman & Palmer, 1996). Basically, assessment collects information using scientific methods systematically and continuously based on a specific context such as a language learning program, a course syllabus, a

language learning theory, existing research, or commonly accepted procedures (Bachman & Palmer, 1996). There are also others who claim that assessment has three main aims: improving teaching and learning processes, motivating students, and providing feedback to students, teachers, and program developers (İnal & Kazazoğlu, 2021; Trotter, 2006). In order to achieve these aims, it should follow certain fundamental principles such as having clear specifications of what is being assessed, identifying and using relevant assessment methods for the purpose of assessment, using a combination of data collecting methods properly, and assessing individuals to make meaningful decisions to function effectively (Linn & Miller, 2005).

As stated before, assessment is the process of collecting information about individuals to make decisions and is important for many reasons. The information gathered can be used in a summative approach to make decisions at the end of the education process to find out how effective the instruction has been or if the individuals have achieved their goals (Göktürk Sağlam, n.d.; Hughes, 1989) as in an achievement test or an end-of-term grade and it may not focus on improving learning and teaching process (İnal & Kazazoğlu, 2021). Thus, summative assessment happens at the end of the teaching and learning process and might be called the "assessment of learning". On the other hand, formative assessment happens during the teaching and learning process, and it is conducted by the teachers to gain insight into students' ongoing learning process and collects information on the learner performance in a systematic and routine way. With formative assessment techniques such as quizzes, peer or self-assessment techniques, in-class observations or portfolios, teachers can check their students' progress, see how much of the learning objectives have been met, and change their future teaching methods and plans accordingly (Hughes, 1989). For this reason, formative assessment can be referred to as "assessment for learning" (Brown, 2003).

In terms of its manner, assessment can be carried out formally or informally. Informal assessment refers to in-class assessment during instruction that is unofficial in nature such as a good comment on student performance or competence, a smiling face put on a student paper or a positive remark on students' answer, which can be unplanned and incidental (Brown, 2003). On the other hand, formal assessments are planned assessment techniques that sample specific behavior from a set of skills and knowledge as in the use

of tests, presentations, or quizzes, all of which are utilized to analyze student achievement (Brown, 2003).

Measurement and evaluation are related to the endpoint of the assessment cycle. Getting the numerical data at the end of an evaluation cycle and agreeing on how much the individual has achieved is the measurement process (Linn & Miller, 2005). In other words, measurement is the quantitative description of an individual's performance, which is always expressed with numerical data. Evaluation, on the other hand, is related to the use of the assessment cycle to make a meaningful decision concerning the teaching and learning program, learners themselves, institutions, or any other domain-related aspect (Bachman & Palmer, 1996). Shortly, evaluation is the process of going over the whole assessment cycle to decide for the benefit of different stakeholders. Thus, it is related not only to test results, but also to the curriculum, program, teaching methods, the teachers, students, and their motivation, resources, staff development and the institution (Richards, 2001).

2.1.1.2 Testing

Language tests are one form of assessment that measure individuals' ability, knowledge, or performance in a specific target language use domain (Brown, 2003). They act as instruments assessing individuals' language skills and abilities in a systematic way by giving them a set of questions to answer or tasks to carry out (Linn & Miller, 2005). They differ from assessments in the sense that they are carried out in a given time frame, with a limited number of questions or tasks, and they collect information about the individuals in one episode of their whole learning process (Green, 2014). Thus, while assessment is a continuous process; testing is assessing learners in a single set time frame.

Tests follow explicit procedures formed in order to get samples of specific behavior, which in the EFL context is the specific language abilities of individuals that the tester is interested in (Bachman, 1990). Good language tests should have certain features, which are: being based on a syllabus, curriculum, or theory that the test takers are aware of, standard in the sense that it provides a good comparison of achievement among different test takers, giving a clear overview of test takers' progress, achievement or strengths, and weaknesses in order to lead to meaningful, educated decisions regarding

students, teachers, the teaching and learning process or research purposes on many different aspects of language (Bachman, 1990; Douglas, 2010). Therefore, a test should use a certain method to accurately measure individuals' performance and linguistic competence about language in a given language domain (Brown, 2003). In an ideal context, course curriculum should clearly specify the learning outcomes and provide a clear measure for students to be assessed and, thus, should contribute to both student learning positively by providing meaningful feedback on their learning and program development by identifying short and long term goals for the future learning environment and finding out ways to overcome future difficulties (Linn & Miller, 2005).

2.1.1.3. Washback

The term washback (backwash also) is defined as the effect or impact of assessment on teaching and learning (Bachman & Palmer, 2010; Hughes, 1989). The assessment that is used will have a natural effect not only on teachers and students, but also on the different parties and institutions involved in the process of assessment such as language learning programs, program developers, and even education systems and the whole community (Bachman & Palmer, 2010). Because these different parties are affected by the decision-making process and the function of assessment, it is inevitable that the assessment affects how teachers teach, how students learn, and their attitudes toward the course content and the teaching program. Thus, due to its multifaceted and complicated area of influence, washback should be closely examined.

Because of the washback of assessment, teachers might find themselves in situations rather incompatible with their regular behaviors. In other words, they might employ methods and strategies they normally would not to prepare their students for the exam, which many call teaching to the test (Bachman & Palmer, 2010). When teachers do that, and the test is either irrelevant to or distant from the course content, the washback happens to be negative as course content is not sufficiently covered so as to prepare the students for the test. However, when teaching aims and test objectives match, the effect of the test on teaching turns out to be positive. A very similar situation is valid for learners as well. When the assessment affects the way learners learn, as in a situation where learners do not get involved in important aspects of teaching which they believe will not be covered in the assessment, harmful washback of assessment can occur. When the test

content and assessment content highly match and learners pay more attention to teaching for the sake of the test, a positive washback is apparent to be present.

Many factors have been assumed to be affecting washback. Teacher-related factors include their beliefs and attitudes toward the exam, textbooks, students, teaching practices, experience and training, and personality (Spratt, 2005). There are also exam-related factors: the stakes of the exam, its purpose, content, and the time when the exam is administered (Spratt, 2005). There are also school and resource-related factors which are the school administration, the school atmosphere, cultural factors and traditions, and available resources at hand (Spratt, 2005).

2.1.1.4. Types Of Tests

Since the main aim of tests is to make educational decisions, tests are grouped and categorized according to their purposes for decision-making in literature. Yet, many tests serve more than one purpose and thus might be difficult to be grouped under certain categories, which does not constitute a problem in terms of testing or referring purposes (Carr, 2011). Below, classifications and categorizations of test types are given by various scholars according to different methods of classification. In this study, Bachman (1990)'s classification according to tests' intended use, content, and frame of reference will be used and explained in detail. Other classifications are also given in the tables below as a summary.

Table 1. Test Types (Bachman, 1990)

Bachman (1990) Classification	
Intended Use	Placement Tests
	Diagnostic Tests
	Achievement Tests
	Proficiency Tests
	Aptitude Tests
Content	Theory Based Tests
	Syllabus Based Tests
Frame of Reference	Norm-Referenced Tests
	Criterion-Referenced Tests

Table 2. Contrasting Categories of EFL/ESL Tests (Madsen, 1983)

Knowledge Tests	Performance of Skills Tests
Subjective Tests	Objective Tests
Productive Tests	Receptive Tests
Language Subskill Tests	Communication Skills Tests
Norm-Referenced Tests	Criterion-Referenced Tests
Discrete Point Tests	Integrative tests
Proficiency Tests	Achievement Tests

Table 3. Classification by Linn & Miller (2005)

Basis For Classification	Type of Assessment
Nature of Assessment	Maximum Performance
	Typical Performance
Form of Assessment	Fixed-Choice Test
	Complex Performance Assessment
Use in Classroom Instruction	Placement
	Diagnostic
	Formative
	Summative
Method of Interpreting Results	Criterion-Referenced
	Norm-Referenced

Table 4. Classification by Tunaboylu (2021)

Purpose	Placement
	Diagnostic
	Proficiency
	Achievement
Function	Summative
	Formative

	Criterion-referenced
	Norm-referenced
Construction	Discrete-point
	Integrative
	Direct
	Indirect

Table 5. Classification by Carr (2011)

Decision-Based Classification	Curriculum Related	Admission, placement, diagnostic, progress, and achievement tests
	Selection Related	Proficiency and Screening Tests
Framework for Interpretation of Results	Norm-Referenced and Criterion Referenced Tests	
	Summative and Formative Assessment	
Other Types	Objective and Subjective Tests Direct and Indirect Testing Discrete and Integrative Testing	
Performance Assessment	Task Completion tests	
	Language Use Tests	

Bachman (1990) classifies tests according to their intended use, the content of the texts, and the frame of reference. Tests according to their use are placement tests, diagnostic tests, achievement tests, proficiency tests, and aptitude tests. In terms of their content, they can be grouped as theory-based language tests and syllabus-based language tests. As for the frame of reference, they can be classified as norm and criterion-referenced tests.

Placement tests are usually used in order to determine the individuals' level or performance before instruction begins and focuses on a set of inquiries such as having enough knowledge and skills to be placed at a level to healthily survive in the planned instruction or to identify the study habits or personality characters of individuals to see

what mode of instruction would suit the individuals better and to make sure that the levels of all individuals placed in one level fit the instruction well enough (Linn & Miller, 2005; Tunaboylu, 2021). The main goal of the placement test is to make sure all the individuals in one level meet the certain requirements of one level of instruction without learning difficulties (Bachman & Palmer, 1996). It should be noted that each institution might have its different needs and expectations of a level; therefore, it may not be suitable for institutions to use ready-made published placement tests but to administer their in-house placement tests to put students into the right levels, which makes placement tests context-bound (Tunaboylu, 2021).

On the other hand, a diagnostic test is used to diagnose an individual's strengths and weaknesses in terms of language ability. They can be used at the beginning of instruction to see what is missing or present in the individual's language proficiency together with the placement tests or separately after seeing each individual is placed in the right level (Carr, 2011). Before instruction begins, they might give the teacher an idea about what to cover in class, which aspects need remedy or extra practice. During the course, they help to see what difficulties individuals experience, possible gaps in instruction, and missing skills in a course to provide a remedy for (Linn & Miller, 2005).

Another test type classified according to its function; achievement tests, which are structured according to a set of objectives, instruction programs, or curriculum, measure individuals' progress over the course (Madsen, 1983; Tunaboylu, 2021). There are two important features of achievement tests. First of all, they need to be related to instruction and based on a certain curriculum. This way, they support instruction by helping teachers and learners see if learning objectives have been met or not. Secondly, achievement tests can be summative or formative in nature depending on when in the course and for what purpose they are carried out. If they are delivered at the end of the course to see if the syllabus has been successfully covered, and the learners have achieved their learning goals, they are called final achievement tests; whereas if they are carried out during the course to see how much learners are progressing or to see the gains or growth of learners in time, they are called progress achievement tests (Tunaboylu, 2021). It must be noted that although achievement tests give a picture of individuals' mastery level of English,

they only compare them against a set of objectives such as a syllabus or curriculum; thus, they do not give a picture of the overall language proficiency level of individuals.

Proficiency tests assess individuals' general level of language in various language skills and components without a set curriculum, disregarding the individuals' previous language learning experience or training (Carr, 2011; Tunaboylu, 2021). They aim to see the readiness of individuals to use the language in a target language domain or identify if the individuals have sufficient command of language to fit in a particular context (Madsen, 1983; Tunaboylu, 2021). Proficiency tests have certain features such as not being curriculum-bound but using generally accepted standard descriptions of language skills, always being summative and therefore not being useful for the classroom setting, and assessing individuals on a wide range of skills through various test methods.

The final test type, according to the classification regarding their function, is aptitude tests. Aptitude tests are used to determine individuals' future success in learning a foreign language as an indicator of language learning potential (Madsen, 1983). Generally, aptitude tests are administered before the instruction begins to determine how ready the learners are for instruction and might be a good method to understand what type of instruction learners might enjoy, what learning problems they might encounter, or how classroom dynamics should be set (Linn & Miller, 2005). It should be noted that aptitude tests do not measure individuals' set capacity to learn a language but rather determine what abilities they possess and how these can be useful for them in future instruction (Linn & Miller, 2005).

When tests are classified according to the content they use, they can be either theory-based or syllabus based. When the test construct is based on a pre-determined curriculum, syllabus, a set of objectives, or a learning program, the language use domain is certain; thus, tests assess if these objectives or aims are achieved as in progress achievement or final achievement tests. The language domain will be specific to the learning context; therefore, individuals will be responsible for those learning elements in the assessment process. As said, achievement tests as well as context-specific diagnostic tests or placement tests with certain pre-set standards can be examples of syllabus-based tests (Bachman, 1990).

On the other hand, theory-based tests are usually proficiency tests, which do not rely on a specific curriculum or a set of objectives but measure individuals' general language mastery based on a language theory (Bachman, 1990). The theories or approaches behind the acquisition of the skills or the components that make up the content of the language test will determine the nature of the language test to be used. However, in some cases, the nature of the theory-based language tests might be similar to those of a syllabus-based test, when the same theory is the root of both assessment and instruction as in a grammatical competence proficiency test and a progress test (Bachman, 1990).

As for the test types according to their frame of reference, it can be said that a norm-referenced test uses a framework where a test taker's score is evaluated in relation to other test takers' results interpreting the individual score according to a mean (the average), median (the middle score), standard deviation (the amount of variance within scores) and a percentile ranking with the purpose of rank-ordering the test takers (Brown, 2003; McNamara, 2000; Tunaboylu, 2021). The idealized distribution of the results reveals a normal-distribution curve where the highest and lowest frequencies of scores are at distant sides of the curve and the largest proportion of scores stay in the middle, forming a bell shape (McNamara, 2000) and thus, an individual's score is interpreted by its position within the distribution range. Using norm-referenced scoring has two main advantages. First of all, it provides quick and easy scoring when the responses to the questions are fixed and determined beforehand (Brown, 2003; McNamara, 2000), which contributes to practicality. When expenses and time are the main concerns, this type of scoring proves efficiency. Secondly, a variety of skills and language components can be measured easily, quickly, and reliably since the test is used more than once in most cases. Piloting, analyzing, and revising ensures the quality and practicality of the test. However, it might be disadvantageous for students to be compared to other students' results as it creates a competitive environment and might be discouraging for average or low achievers (McNamara, 2000).

Criterion-referenced tests, however, measure individuals against verbal descriptions of language capabilities, enabling instructors to see if learning objectives are achieved (Brown, 2003; McNamara, 2000). It allows seeing how much individuals can achieve according to their age, career, or level (Tunaboylu, 2021). After the assessment

process, the students are usually given feedback in grade or report format. As the students' rank within a range is not important, it is assumed that criterion-referenced testing increases student motivation by encouraging them to do their best personally (McNamara, 2000). This type of testing has advantages such as easy measurement of learning objectives, being a good indicator of learning, and easing the decision-making processes by guiding the teaching and learning process. The disadvantages might be being time and effort-consuming to create because of being tied to a curriculum (Brown, 2003) and the difficulty of setting appropriate criteria to evaluate learning.

Different scholars also use other types of classifications according to their function, purpose, content, or method. Discrete point testing and integrative testing are associated with the methods of testing and the test content. Direct and indirect testing is related to the methods used in language testing.

Discrete type of testing has been influenced by structuralist linguistics (McNamara, 2000) and bases its roots in the understanding that language can be broken into its segments such as skills and components, and these different segments of language can be decontextualized and tested individually and independently as each language element is tested with a different item (Brown, 2003; Hughes, 1989). According to this understanding, there can be separate reading, listening, speaking, and writing tests as well as assessments that test individuals' grammatical, syntactical, morphological, and phonological competence, presented in an isolated way without much context. In order to have quality tests, there should be as many items as possible to test all skills and components according to discrete point testing arguments.

Integrative testing is based on the understanding that language elements cannot be divided from each other, and test takers should be assessed from all aspects of language (Brown, 2003; Tunaboylu, 2021). According to integrative testing, communicative competence is a wholesome skill, so language must be integrated with all aspects and skills. Thus, language elements and skills can be tested together while completing a task such as a cloze test or a dictation activity (Brown, 2003; Hughes, 1989).

On the other hand, direct testing can be explained as directly assessing what skills need to be assessed. In direct tests, test constructs are certain, test tasks are particular to

the pre-determined construct, and the skill is directly observable in the test (Fulcher & Davidson, 2007). This type of testing is advantageous for the teacher as the test conditions and construct can be easily created, learner performance can be easily observed, and it is possible to test productive skills. In contrast, an indirect test does not require direct testing of the skills or components. Similarly, the results do not need to be observed directly but indirectly, as in, for instance, testing the individual's writing skills through an error correction test (Fulcher & Davidson, 2007). The advantages of indirect tests can be representing real-life situations better and presenting a better overview of the individuals' language use skills as the test takers are required to fulfill one task to show performance on a different construct. In literature today, all language tests are considered indirect measures as the test constructs are not directly observable but require interpretation of results (Fulcher & Davidson, 2007).

2.1.1.5. Aspects of Language Assessment

For a test to be effective and dependable, it has to confirm three key qualities: reliability, validity, and practicality. In order to make healthy decisions about students, instruction, learning, and institution, solid knowledge of test qualities is important. Three main qualities of good language tests and some other important qualities are explained in this subsection.

- **Reliability**

Reliability refers to consistency in measurement. In other words, it can be explained as the generalizability of test scores at different times, with different scorers, or with different sample behaviors of the same domain tested (Linn & Miller, 2005) under the same conditions (Madsen, 1983) and this safe reproduction of the scores is referred to as reliability. As reliability is related to the scores of a test, it is mainly concerned with numerical data, and thus, analysis of reliability is possible. Since assessments are used to make decisions about individuals' learning in many different aspects, the reliability of test scores is a highly desirable feature for educated decision-making.

It should be noted that perfect reliability is not attainable for tests (Carr, 2011; Farhady, 2021; Linn & Miller, 2005) since tests are bound to many different variables such as the physical and mental condition of the test taker, memory differences between

time intervals, guessing factor and so on. The slight differences between the scores given at different times indicate the error of measurement and can be simply explained as the difference between the observed score and the individual's true score (Farhady, 2021). Yet, as the main aim of the assessment is to accurately measure individuals' performance, the closer both scores are to each other, the more reliable assessment becomes. Yet, an individual's true score can only be calculated by testing the same individual over and over again with the same assessment material assuming that the test is perfectly reliable and all the conditions regarding the exam and the individual are exactly the same, and the time interval does not affect any variables; which is almost impossible (Hughes, 1989). Because of this, reliability has been calculated through various methods, all of which rely on correlating two sets of scores with the same or similar procedures in different ways such as *the test-retest method, equivalent forms, split half method, Kuder- Richardson, Coefficient Alpha, and interrater reliability estimating methods* (Linn & Miller, 2005). All of those methods yield to a correlation coefficient score that can take a value between 0 and 1; 0 means no reliability, and 1 is the highest reliability, yet both extreme values are generally unseen in practice (Carr, 2011; Linn & Miller, 2005). The more important the decisions are taken according to the test result, and the higher the stakes become for a test, the higher the correlation coefficient is sough (Hughes, 1989). In other situations where a high-reliability level is necessary, the decisions to be made according to the test result are final, irreversible, unconfirmable, and have long-lasting consequences (Linn & Miller, 2005).

To ensure reliability, the test-retest method can be used. In this method of ensuring reliability, the same test or assessment can be given to the same group of students twice, generally within one or two-week period of time intervals, to prevent students from remembering the questions (Farhady, 2021). The scores of both tests are correlated, and the correlation coefficient is obtained at the end of the calculation process, indicating the level of stability for the test. Thus, this method is also called as the measure of stability (Linn & Miller, 2005). Despite its advantages such as being easy to administer, cost-saving, and reliable as a reliability calculation method, it has some shortcomings. First of all, the time interval between the two tests might affect the reliability negatively as when the time interval is short, test takers might remember their responses easily, and when the

time interval is long, the test takers might add to their existing knowledge of the subject matter, both of which will decrease stability and reliability significantly (Farhady, 2021; Linn & Miller, 2005). Secondly, test takers might be unwilling to take the same test a second time, or it might be meaningless and time-consuming for the class teachers to administer it again with the same group of students. Because of this unwillingness in both parties, the test-retest method may not provide a clear picture of reliability.

Another method of ensuring reliability is the equivalent or parallel forms method. To calculate reliability through this method, two similar forms of a test, in terms of both test content and difficulty level, are given to the same group of students in close time intervals, and the scores of two tests are correlated to reach a correlation coefficient value (Linn & Miller, 2005). It is based on the assumption that both tests measure the same set of behaviors, and the time interval does not cause any changes for students. It is advantageous over the test-retest method as it eliminates the impracticality of applying the same test to students. However, it also has shortcomings as it is highly likely that two tests will not measure the same sample of behaviors, decreasing reliability. Moreover, similar to the test-retest method, the time interval might negatively affect the individual's performance. When the time interval is too long, the level of knowledge individuals might change within this time (Linn & Miller, 2005).

The reliability coefficient can also be easily measured with the split-half method, which is done when the test developer divides the test into two subsets of questions before test administration, each subset representing the same assessment. Then, the test is delivered to a group of individuals. This way, each student is provided with two scores representing the same set of questions, and when correlated, the correlation coefficient value is easily obtained (Linn & Miller, 2005). This method might provide a more reliable score for the reliability coefficient since neither the same test is administered twice nor is there a time interval between the administration of two sets of questions (Linn & Miller, 2005).

Other methods such as the Kuder-Richardson formula and the Coefficient Alpha allow test developers to calculate reliability with a single administration of the test. With the Kuder-Richardson formula, the test is not divided into equal halves; but a simple

formula calculates the mean of all possible split halves in the test set (Farhady, 2021). This type of reliability measurement is only applicable to dichotomous items whose score is either 1 or 0 (Linn & Miller, 2005). Another method that parallels all possible split halves in a test is the popular Cronbach's Alpha reliability measure, which depends on the pre-acceptance of two features: homogeneity and item independence (Carr, 2011). The reliability score will be inaccurate when the items are not parallel or homogenous. In addition, Cronbach's Alpha treats the items as individually standing items, not being related to each other. In other words, the performance of one item does not affect the performance of another (Carr, 2011). Since both of these methods check reliability on one set of questions, they check the internal consistency of a test rather than checking the stability of reliability over time (Linn & Miller, 2005).

As for subjective testing, reliability can be assured through rater reliability. When the scores of an assessment are not based on right or wrong answers, in other words, when the scoring of a test is not objective, it is necessary to ensure rater consistency and reliability. It is highly important to check rater reliability to guarantee fairness among student scores, which prevents some students from getting very high or low grades against a set of criteria because of rater leniency or strictness (Linn & Miller, 2005). In other words, interrater reliability makes sure that very similar scores are given to students regardless of who is grading and when they are grading with the help of pre-set criteria and rater training activities. The rater consistency is two-dimensional as the raters should be consistent with their own grades on the same test on different marking sessions, and it is also necessary for two or more raters to grade the same test in a consistent way (Farhady, 2021).

- **Validity**

Another significant feature of a test is validity. In literature, a test is considered valid if it measures what it has to measure (Hughes, 1989; Madsen, 1983). In this sense, a simple explanation in the EFL context would be the expectation of language-measuring questions in a language test without involving any other irrelevant elements or characteristics. Thus, if the test content consists of items that indicate a good representation of the target language domain, the test is said to be valid. However, this

definition and explanation have been limited to only content validity, and has been expanded to other aspects in time (Carr, 2011; Linn & Miller, 2005) due to the fact that tests should not only give evidence for content representation but also show that it leads to an adequate and appropriate future reference about test scores. Thus, the understanding that the validity of the test is limited to its content shifted to a more comprehensive understanding, which made test validation an ongoing process of collecting scientific evidence about examination, analysis, and use of test results (Uysal, 2021). This process involves providing explanations on purpose, learner characteristics, difficulty level, the content of the test, the language theory behind the test, text types and methods used in the test, and criteria for the grading and scoring of the test (Alderson, 1995 as cited in Uysal, 2021).

According to Carr (2011), there are some important points to consider concerning validity. First of all, validity is related to how appropriate and useful test results are in terms of performance interpretation, not the procedure itself although the term is almost always used as test validity (Carr, 2011). Secondly, validity is almost never inexistent or totally present in an assessment, but it is rather explained in varying degrees, which means that an assessment is neither valid nor invalid (Carr, 2011). Therefore, an assessment can be expressed as highly valid, moderately valid or having low validity. Third, validity is specific for each test, purpose, and condition; thus, no assessment task is valid for all conditions and purposes, and it is important to consider these purposes and conditions while explaining or measuring validity (Carr, 2011).

There are different aspects of validity that explain different validity issues with different methods. Although some of these concepts may have evolved in time, most of them are still valid for validity today. Thus, to understand the validity degree of a test, it is important to refer to the following points regarding validity. There are a variety of validity types. These are face validity, content validity, construct validity, response validity, criterion validity, predictive validity and consequential validity.

Face validity answers the questions of clarity, relevance, and appropriateness of the test for the intended goals (Uysal, 2021). A test is considered to have face validity as long as it looks as if it measures what it is supposed to measure. Face validity is not

measured empirically, thus, an analysis of face validity is rather subjective (Hughes, 1989). At this point, the judgment of teachers, experts, or test takers should be taken into consideration to decide if a test possesses face validity. It is one of the integral validity aspects of a test; however, face validity has been criticized by many for leading to poor judgment about the test by creating test appeal, but not effectively measuring what it should be measuring and affecting the decision-making purpose of assessment negatively (Bachman, 1990).

Content validity is related to the internal validity of the test and answers the questions regarding the relevance of the test content to the target language domain content and the intended purpose (Uysal, 2021). In other words, it means if the skills sampled in the test represent the domain of skills to be evaluated and if it samples the most important focus of the target language content well enough (Linn & Miller, 2005). The most important aspect of content validity is representing and including all the significant points in the course objectives or curriculum. Thus, a good and sound knowledge of the course curriculum, target skills, structures, and content is required to measure if an assessment has content validity. A test with high content validity proves to be successfully measuring student ability, and the use of the scores indicates good judgment. Thus, a test that does not fully or accurately represent course content would result in inaccurate results and might negatively affect instruction (Hughes, 1989). In sum, content validity is measured by comparing test specifications or items with the course objectives by teachers or experts and is based on logical evaluation of the correlation between the two (Hughes, 1989; Linn & Miller, 2005; Uysal, 2021).

Another important point in validity is construct validity. Construct validity, which is an aspect of internal validity, answers the questions of how well the test content represents target language domain tasks or how much the test content is relevant to the goals of the assessment (Uysal, 2021). In other words, it shows how much the performance shown in the test matches the characteristic or quality that the assessment intends to measure (Linn & Miller, 2005). In order to ensure construct validity, a test needs to correlate to some extent with other tests that theoretically measure the same skill or language aspect and should not correlate with tests that suggest measuring different skills or aspects. If it does, it shows that the test also measures another construct and thus proves

to have low construct validity. To measure construct validity, the test items or tasks should be examined and correlated with the specific language teaching theory to teach that specific aspect (Uysal, 2021). Also, a statistical analysis such as factor analysis or convergent-divergent validation methods can be used to measure construct validity, but it cannot be measured precisely or expressed statistically (Carr, 2011).

Score or response validity, which is an integral validity aspect, basically refers to the items of the test. It analyzes the variety of items, response formats, discreteness or integrativeness of the items, and if the items assess the same objectives in different ways (Uysal, 2021). To ensure response validity, a statistical analysis such as item analysis, internal correlation analysis, or rater reliability analysis should be done (Weir, 2005 as cited in Uysal, 2021).

Criterion validity, which is an external validity aspect, analyzes how well the test scores correlate with that of another trustworthy assessment (Hughes, 1989; Uysal, 2021). At this point, the analysis of both tests should sensibly relate to each other and show a meaningful correlation to ensure high criterion validity. There are two types of criterion validity; concurrent validity and predictive validity. In concurrent validity, test scores of a represented coverage are compared to the test scores of a full content test. If the scores correlate to a high degree, the test is considered to have a high concurrent validity measure. At this point, the full correlation would take a validity coefficient score of 1, and no correlation would take a score of 0, showing no concurrent validity score for the represented coverage. It should be noted that concurrent validity is not necessarily measured against a longer test, but a meaningful test by the teacher can also suffice (Hughes, 1989).

Predictive validity refers to the degree which shows how well a test can predict an individual's future performance. It simply correlates the scores of an individual on a test with a future test, which gives a correlation coefficient value for the two. It should be noted that predictive validity can be affected by many outer variables such as present and future learning, test takers' emotional or physical well-being, and test conditions (Hughes, 1989).

A rather new concept, consequential validity, which is an external validity aspect, answers the questions regarding the scores of the test as in, if the implications of the score interpretation have a scientific basis, if the long and short-term results of the test scores support the test aims or not, and if there are any negative washback effects of the test on teaching or on the curriculum (Uysal, 2021). As the assessment process highly affects decisions regarding teaching and learning in general, consequential validity has recently gained great importance for the stakeholders (Uysal, 2021). Thus, it has become paramount to evaluate the consequential validity of the assessments by collecting data through surveys, questionnaires, interviews with test takers and makers, or classroom observations after assessment.

- **Practicality**

The final important feature of an effective language assessment is practicality. It is generally described as the correlation between the resources needed for assessment and resources at hand. If the available resources are more plentiful than the resources necessary, then the assessment seems practical; whereas if the necessary resources exceed the available resources, the assessment is impractical (Bachman & Palmer, 2010). If the assessment proves to be impractical, no matter how well it assesses the individuals, it might be postponed to use, may not be used after a few times, or might be canceled without being used. Thus, practicality is one of the most important qualities of a test, even if it does not entail a statistical analysis or yield a numerical value.

A test can be called practical if it is cost-effective, time-saving, easily administered, and scored. The practicality of an assessment is necessary for all phases of test development and administration. Thus, test developers should consider practicality in planning, developing the test, administration, scoring, and result announcement (Kömür & Erol, 2021). At this point, it can be said that practicality involves aspects of human resources, time, and materials (Bachman & Palmer, 1996). It should be noted that practicality is institutional and context-bound; thus, a test that is practical for one context may not be practical for another (Kömür & Erol, 2021).

2.1.2. Testing Theories for Language Assessment

Like all areas of education requiring testing, assessment and measurement, language education necessitates measurement methods that work well to make meaningful decisions. Thus, issues generalizable to all educational areas such as choosing the right method of measurement, collecting a relatively good and meaningful behavior sample from learners, and defining constructs appear to be worth identifying in language testing, assessment and measurement (Szabó, 2008). Many of these issues have been well addressed by the introduction of communicative language teaching in general; however, some practicality-related issues have been handled by either classical test theory or item response theory, or a combination of both theories (Szabó, 2008). The following section will summarize the classical test theory and item response theory.

2.1.2.1 Classical Test Theory

The Classical Test Theory measures an individual by making assumptions about the relationships between the individual's true and observed score taking the measurement error into consideration (Fulcher & Davidson, 2012; Szabó, 2008). The theory suggests that the true score of the individual is the observed score of the individual added on the measurement error which refers to all possible other influences on the test taker such as variables ranging from exam administration conditions to scoring procedures that might distort the exam score, affecting the test taker's actual score (Fulcher & Davidson, 2007). In this sense, the measurement error should be as small as possible to increase the reliability of the measurement tool. However, in language testing, both observed score and measurement error might change and fluctuate under different circumstances; thus, the reliability of the exam might be challenged. At this point, test reliability can be ensured by some statistical methods such as the test/retest method, parallel forms, split half method, Kuder-Richardson, or Coefficient Alpha methods (Szabó, 2008) which are mentioned earlier in detail.

While analyzing a test against classical test theory principles, it is important to discuss an item, namely a question/ task that requires the individual to present a language skill, in terms of its difficulty and discrimination levels. Item difficulty or facility refers to how difficult an item is and is calculated by taking the average score of all individuals who took the test for each specific item. In other words, item facility is calculated by

dividing the number of test takers who answered that specific item correctly by the total number of test takers. After this procedure, an item facility value between 0 and 1 is obtained. The score is interpreted as the closer the value is to 1, the easier the item is, and vice versa for 0. The desired value varies according to institutional expectations, test type, test purpose, and the time of testing. Item discrimination refers to how well an item discriminates between high and low scorers in a test. It is obtained by subtracting the average score of the low achievers in the total test-taker population from the average score of the high achiever group. This calculation gives an item discrimination value between 0 and 1, and the higher the score gets, the better discriminative the item becomes. Although these statistical analyses give valuable information about and insight into single items, they must not be considered as the sole and ultimate sources to determine the reliability of items as many other variables must be taken into consideration while making decisions because, for instance, different test taker groups may yield to completely different item analysis results (Szabó, 2008). Besides, these statistics do not give proper data while analyzing subjectively marked assessments. Moreover, this type of test theory provides information on only one source of measurement error, which is the item without further analyzing other factors (Szabó, 2008). Another issue with classical test theory is that the reliability and validity of tests are highly affected by the extremes in both the high and low-achiever sides of the continuum (Brown, 2012). For these reasons, item response theory might provide an additional viewpoint on the language testing theory to classical test theory.

2.1.2.2 Item Response Theory

Item response theory (IRT) is also known as the latent trait theory and refers to using observable assessments to measure unobservable traits of an individual test taker. In other words, it means using measurable skills to measure a skill normally hidden within the total skills of an individual test taker as in, for instance, assessing an individual's reading skills through a multiple choice test given after a reading passage (Ockey, 2012). Each test item in IRT presumably has a certain, set level of difficulty, and by putting each item on a scale of difficulty, it in fact helps decision-makers determine the level of success for each individual according to their achievement of items on the continuum (Aydın, 2004). In the IRT model, each item is evaluated separately, meaning each accurately

answered question might reflect the achievement on a single separate skill (Ockey, 2012). In short, the IRT model assumes that high achievers will be more likely to answer more difficult questions correctly and low achievers will answer those questions mostly incorrectly, yet there could also be students who incorrectly fit into one category (McNamara, 2000; Szabó, 2008).

IRT has many advantages. First of all, as each item is measured separately, assessment becomes independent of the test taker population. Other advantages are facilitating the production of tests with equal difficulty levels, providing a great deal of information on various aspects of tests such as raters, items, and test takers, provides test linking which refers to the process of using tests with calculated difficulty levels to observe progress over time. Furthermore, it allows item banking by storing each item separately and creating opportunities for adaptive computer tests (Aydın, 2004; McNamara, 2000; Ockey, 2012).

Despite its advantages, IRT also has some drawbacks. First of all, the lack of expertise in using computers and software enabling test or item writers to use IRT for testing is limited, so item writers are unwilling to use IRT statistics (Ockey, 2012). Secondly, a large student sample is necessary to apply IRT which might be difficult in classroom settings (Aydın, 2004). A final problem could be the low-achieving students answering more difficult questions correctly as it might be difficult to interpret if this misfitting is because of the individual test taker, the item itself, or because of another factor (Szabó, 2008).

2.1.3. Online Testing and Assessment

Widespread use of computers and rapid development of IT have led to important changes in both learning and assessment, paving the way for the implementation of online learning (can also be referred to as digital learning or e-learning) activities and experiences and online assessment (digital or e-assessment) procedures (Çağlayan, 2021; Linn & Miller, 2005; McNamara, 2000). Online assessment refers to all assessment-related activities, including design and delivery of assessments, marking activities regardless of being done with online tools or humans, storage and transfer of data and results, and is carried out with the use of digital technologies such as computers or online platforms

(Joint Information Systems Committee, 2007). In short, online assessment can be defined as all types and activities of assessment that have an online component or task.

Early digital assessment practices were in the form of projects and certain tests administered merely in university courses (Chapelle & Voss, 2008) and many online assessment tasks were part of formative assessment in institutions in the format of projects and e-portfolios. However, with the increasing ease and convenience of the Internet, many forms of assessment started to be used in various institutions and by internationally recognized assessment programs such as TOEFL in the form of computer-based testing (McNamara, 2000) and Internet-based testing and PTE Academic Online. Thus, as a result of the latest developments and research, it can be said that online assessment is currently used for diagnostic, summative, and formative assessment purposes (Kundu & Bej, 2021). Similar to traditional assessments, online assessments have to follow basic principles of testing which are validity by assuring the selection and application of related and authentic content and items, reliability by giving consistent and transparent results, practicality by being easy to administer, score and store and having reasonable financial and human-related resource requirements and security by being free of cheating or other misconducts (Kundu & Bej, 2021). It can be conducted synchronously when students are asked to complete the given tasks in real-time, under the supervision of the invigilators or teachers, or asynchronously when the students allocate their own time to complete the tasks (Çağlayan, 2021).

There are many reasons behind the increasing use of online assessment in many aspects of learning and in various levels of instruction. The main reason is the increasing use of technology in instructional activities and course books that are being used in instruction (Boitshwarelo et al., 2017). Nowadays, technology is involved in almost all language teaching coursebooks in inputs or in question banks that are automatically provided with the books. When technology is widely used in education, assessment also needs to adapt to the change. Moreover, the student numbers have been increasing in many institutions, which necessitates effective and efficient use of resources at hand, including human resources. When the numbers of students enrolled in classes are high, instructors need to find effective ways to grade students and provide feedback in a timely manner, and the use of technology in assessment can cater to this need (Boitshwarelo et al., 2017).

Online assessment can be very effective when some key aspects are taken into consideration (Boyle et al. 2003 as cited in Gaytan & McEwen, 2007). First and foremost, making use of realistic scenarios and tasks and successful alignment of course objectives and assessment scenarios should be considered while assessing students online. Moreover, making use of the assessment software as soon as possible to make learners familiar with the systems and providing IT support and mentoring when needed are also extremely important to provide a reliable online assessment experience for students. Finally, providing learners and instructors with necessary training is vital for a positive online assessment process (Boyle et al. 2003 as cited in Gaytan & McEwen, 2007).

According to the results of a study by Walker (2007), effective online assessment has to follow certain principles which teachers, test takers, and item writers or task providers should consider while planning online assessments.:

1. Assessment should be perfectly aligned with the course curriculum and objectives.

2. Instructions should be clear and precise.

3. The feedback to the assessment should fit with the purpose of the assessment.

4. Marking procedures should be completed transparently.

5. Students need to be informed about the process.

6. Students with special needs should be taken into consideration.

7. There should be a good variety of question types and task types.

8.Tasks and items should be level-appropriate.

9. Timing should be arranged meaningfully both in terms of the length of the assessment and the timing of the assessment within the course.

10. Neither the items nor the tasks should test students' IT skills or computer use expertise.

Apart from summative assessment practices such as diagnostic, achievement, and proficiency tests which might include multiple choice, drag/drop, true/ false or fill-in-the-blanks type of questions (Rolim & Isaias, 2018), online assessment has been innovatively used in scenario role-plays, e-portfolios, self-reflection and self-assessment tools, wikis

and blogs, and collaborative assessment tools (Appiah & van Tonder, 2018; Joint Information Systems Committee, 2007). Teachers are expected to form the tasks, conduct those with the relevant group of students, assess and grade the tasks, and give feedback with the help of digital tools such as computers and the Internet (Çağlayan, 2021).

2.1.3.1. Online Assessment Related Theories

Although it is a field that has started to develop rather recently compared to traditional types of testing and assessment, online assessment bases its roots in both recent and profound classical theories. One important theory that affects online assessment is the Cognitive Load Theory which proposes that the amount of information that can be stored by an individual is a lot less than the amount of information that can be processed by the individual (van Merriënboer & Sweller, 2005). According to this theory, the intrinsic cognitive load is the effort required by the nature of the task that is being presented; the extraneous cognitive load is the extra load that comes with the way the task that is being presented, and germane cognitive load is the mental effort necessary to combine new and existing information to fulfill the task (van Merriënboer & Ayres, 2005). With regard to cognitive load theory, online assessment should minimize the extraneous cognitive load of the task while maximizing the germane cognitive load by improving the validity and reliability of the assessments via asking the individuals to perform focused, clear, well-organized, and clearly defined tasks that require utilizing real-life skills while performing them. This way, the extraneous load of the tasks will be lowered, yet the germane cognitive load will be increased.

Another key theory that has affected online assessment is the Technology Acceptance Model, which suggests that an individual's level of acceptance of a new technology is highly affected by his perceptions of its usefulness and ease of use (Davis, 1985). This theory aims at finding out how and why individuals make use of technology and accept it. Perceived usefulness means that if a person believes that using that specific technology or technological tool will ease their job, he tends to use it and perceived ease of use means how effortless the new technology or technological tool is to the user (Davis, 1985). Therefore, if teachers or students perceive online assessment as a useful and easy-to-use tool, they tend to have a positive attitude towards its use, accept and use it now and in the future. In online assessment, this could be achieved by providing necessary support

and guidance, as well as providing easy-to-use tasks to both students and teachers. In the context of online assessment, the test takers' and teachers' attitude towards the use of it is highly significant as they might have a positive or detrimental impact on the practicality, reliability, and validity of the assessments. If there is an issue regarding the validity or reliability of the assessment, the test results may not reflect the individual's true score (Alkış & Özkan, 2010).

Constructivism is another theory that affects online assessment. According to constructivism, learning is an active process in which learners construct knowledge by using their experiences and interactions with the environment; thus, the learner has an active role in creating meaning through the learning material (Benson & Brack, 2010). Learners both use their existing knowledge and new experiences to build meaning according to their own sense of the world around them. The relationship between constructivism and online assessment suggests that the assessment tasks should allow learners to make meaning themselves utilizing of the task and their existing knowledge in a meaningful way, rather than simply reciting memorized rules or formulas (Benson & Brack, 2010). Thus, learners should engage in meaningful activities that would help them experiment with the language in a productive way such as engaging in writing or speaking tasks. Moreover, constructivism values immediate and personalized feedback to learners to allow self-reflection; therefore, online assessment is highly advantageous in this perspective.

2.1.3.2. Advantages of Online Assessment

Online assessment has certain advantages for stakeholders such as test takers, test developers, teachers, and institutions. One significant advantage for test takers is providing solutions to logistics problems and flexibility by allowing the test takers to take the test in their own time and at their convenience, in the comfort of their place regardless of physical distances (Alruwais et al., 2018; Long et al., 2018). Second, as long as the system has a user-friendly and easy interface, students enjoy being assessed with the help of technology and become more engaged (Alruwais et al., 2018). It also allows for personalized testing and helps increase learner motivation by involving technology use, which greatly appeals to today's learners (Alsadoon, 2017; Alruwais et al., 2018; Ridgway et al., 2004). Also, their responses are recorded and scored easily and quickly, with almost

no errors, especially with objective items. Even subjectively graded exams can be marked easily when correct algorithms or artificial intelligence are provided and utilized (Long et al., 2018). It increases reliability by letting assessment be free of racial, ethnic or social discrimination grounds (Özden, Ertürk & Şanlı, 2004, as cited in Alsadoon, 2017). Finally, it allows immediate feedback for learners so that they can learn from their mistakes in their performances and increase their motivation and engagement (Öz, 2014b).

Online assessment provides advantages for test writers as well. It improves practicality for test developers by allowing them to write and upload the questions anytime and anywhere with the convenience of making last-minute changes to the questions (Long et al., 2018). Moreover, it provides the chance to conduct testing on-demand when there is a need and adapt the test when there are different test purposes (Ridgway et al., 2004). Authenticity is another advantage of online testing as various tasks, items, and input types such as interactive tasks, visuals within listening tasks, or real-life simulations can be used for assessment (Chapelle & Douglas, 2007; Öz, 2014b). Finally, it helps increase practicality by providing a time-saving solution to traditional testing by freeing test developers from the stresses of high costs or long photocopying times (Alsadoon, 2017; Ridgway et al., 2004).

Online assessment has certain advantages for teachers as well. First, it helps teachers to follow their students' success easily by providing them with immediate data on their learners' performances and also by helping them monitor their improvement by providing easy analysis of student performance and improvement (Alsadoon, 2017; Alruwais et al., 2018). This way, teachers save time that is spent on analyzing student performance and can give quality feedback to students in a timely manner. Moreover, it facilitates group work as the physical distance is disregarded and allows teachers to grade student performances in group work in an easier way (Alruwais et al., 2018). Finally, it allows teachers to assess students' different skills by testing higher-order skills such as critiquing, reflection, and problem-solving skills by utilizing different technological tools and applications (Alruwais et al., 2018).

Online assessment also has some benefits for institutions as it allows grading many student performances within a short period and, thus, allows assessment of large groups

of test takers immediately and effectively (Alsadoon, 2017; Alruwais et al., 2018; Appiah & van Tonder, 2018; Dermo, 2009). Second, utilizing technology reduces costs for institutions as it frees the institutions of photocopying expenses (Alruwais et al., 2018; Appiah & van Tonder, 2018). Finally, it provides more accurate results by decreasing cheating chances by providing an ID check and verification system and using the option to ask different test takers the same questions in varying orders (Alruwais et al., 2018).

2.1.3.3. Concerns about Online Assessment

Despite its many benefits, online assessment poses many difficulties to different users of the systems. Some of these challenges concern test developers, and some concern test takers. There are also other challenges related to different stakeholders such as the institutions or parents.

Online assessment has been criticized for creating challenges to test takers for providing little or no contact with instructors or test providers and for having a lack of interaction (Abduh, 2021). When the interaction is limited, this might create stress and tension for the test takers. There could also be some technical problems or difficulties such as not having an easy interface to navigate around or limited technological support both for the learners and teachers (Abduh, 2021). Another issue related to online assessment might be creating an unfair advantage for technologically or economically advantaged students over others. When students have better computer literacy or economic power to afford the technology, they might perform unfairly better than others (Ridgway et al., 2004).

Online assessment might also be challenging for test developers in many aspects. Preparing and controlling questions requiring technological expertise can be difficult and time-consuming (Alsadoon, 2017; Alruwais et al., 2018; Öz, 2014b). Another problem is related to grading subjective assessments as it might be very difficult to set up a system that would score performances reliably. Also, grading items that might require open-ended responses or be partially correct can be difficult with online exams as it might require writing down all alternative answers that might come up as student responses (Alruwais et al., 2018; Öz, 2014b). Another problem is the problem of security, dishonesty, or academic integrity. The problem might appear as the difficulty of test taker identification,

experiencing uncontrollable cheating occurrences, or the system allowing the test taker to take the test multiple times if not pre-arranged by the system to do otherwise (Abduh, 2021; Alsadoon, 2017; Ridgway et al., 2004). This might also occur as the result of dishonest behavior when students claim to have a power cut or technical difficulty during the exam to take it once again (Alsadoon, 2017). Another example of dishonest behavior could be getting help illegally from outsiders or overriding the security measures taken by the test providers (Alsadoon, 2017). Other integrity problems might occur as students copy texts from the Internet or cheat from each other.

Like test developers and takers, institutions might be challenged by online assessment since it creates certain institutional requirements such as setting up the system and providing the necessary infrastructure for the technology (Appiah & van Tonder, 2018; Ridgway et al., 2004). Moreover, the institution must provide staff with expertise and technological support when necessary (Ridgway et al., 2004). In addition, teachers might require technical support or training before the assessment in order to get familiar with the process and procedures and during assessment against system failures or technical problems (Alruwais et al., 2018).

2.1.4. Perspectives Of Online Assessment

In terms of their meaning, the words belief, conception, perception, and perspective are related concepts that are mostly used interchangeably in the literature. However, they differ slightly in meaning, and it is worth mentioning the differences among them to clarify the concepts. Muñoz et al. (2012) describe beliefs as "a mental representation of reality which contains meaning, preferences, and attitudes that allow the rationalization of complex and different categories of experience" (p.144).

Although beliefs are considered as one of the most important factors that shape indviduals' attitudes toward and behaviors about a practice, recent research has shown that beliefs cannot affect teacher behaviors alone, rather classroom practice has been affected by a combination of various factors, namely values, commitment, motivation, task perception and job satisfaction (Ferretti et al., 2021). This brings about the importance of task perception and in the context of assessment; perception of assessment. By dictionary definition, perception and perspective are quite close to each other. Human

beings perceive the world around them by their sensory-motor experiences and perception is the process of receiving and interpreting this sensory-motor experiences to make up a mental image or understanding of the world around them and these sensory-motor experiences are heavily affected by individual ideas, experiences, and interpretations (Brown, 2002). However, perspective is the individuals' personal viewpoint of the situations and shaped by their perceptions of the world around them. Perceptions and conceptions are not knowledge, truth or they are not based on scientific facts, and thus, they might be incomplete, inaccurate, or inexact (Ferretti et al., 2021).

Since the behaviors or attitudes of teachers towards assessment are highly influenced by teachers' perceptions or perspectives of assessment, it is important to devote attention and effort to learning these perceptions and perspectives (Brown, 2002; Ferretti et al., 2021). According to Brown (2002), teachers have significant conceptions about assessment which are assessment being useful for giving information to improve instruction; assessment being necessary for making students and institutions accountable; and assessment being irrelevant to teaching and learning. As can be seen, classroom assessment is highly affected by teacher perceptions and perspectives, and it is very important to identify these perspectives and perceptions.

An individual's intention to behave in a particular way or use a method is highly dependent on the forces, perceptions, perspectives, or drivers that form or affect this behavioral intention, and an individual's intention to utilize a system or method is affected by the individual's attitude, cognition, social norms and the context of possible implementation (Teo & van Schaik, 2012). Thus, an individual's decision about a particular behavior can be explained by Ajzen & Fishbain's Theory of Reasoned Action and Ajzen's Theory of Planned Behavior (Ajzen & Madden, 1986; Madden et al., 1992). The Theory of Reasoned Action suggests that an individual's behavioral intention determines if the person is going to carry out a particular behavior. This intention is determined by two factors which are a person's attitudes toward the behavior and their subjective norms or the perceived social behavior to carry out the behavior (Madden et al., 1992). The Theory of Planned Behavior is the extended model of the Theory of Reasoned Action with an additional factor. Perceived behavioral control is added to the factors leading to behavioral intention. Perceived behavioral control is the belief or

perception about how easy or difficult performing the behavior will be, and according to this version of the theory, a person's behavioral intention is affected by their attitudes, subjective norms, and perceived behavioral control (Ajzen & Madden, 1986; Madden et al., 1992). As one can see from the figures below (Figure 1 and Figure 2), different from the Theory of Reasoned Action, the Theory of Planned Behavior suggests that perceived behavioral control might lead to the behavior itself without going into the phase of behavioral intention. Thus, it is vitally important to identify a teacher's perceptions or perspectives to understand if they will lead to behavioral intentions and behavior itself.

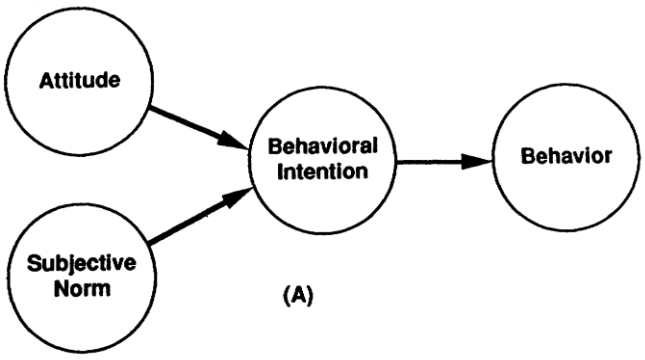

Figure 1. 1 The Theory of Reasoned Action (Madden et al., 1992)

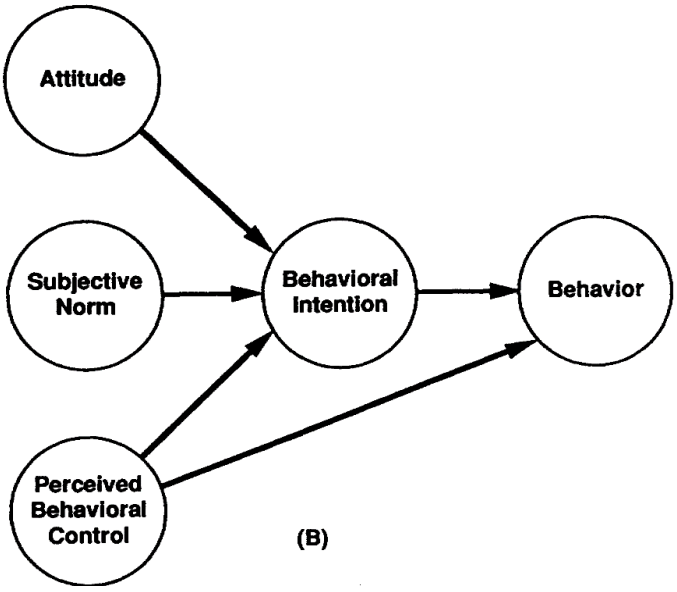

Figure 1. 2 The Theory of Planned Behavior (Madden et al., 1992)

2.2. Research results on Online Assessment Perceptions

Since the prevalence of online learning and teaching with the advancing Internet over the recent years, there has been a need to evaluate students via corresponding methods and technologies (Alghammas, 2020). Therefore, many higher education institutions have adopted methods to evaluate students with the use of information technologies, and this has brought about online assessment methods (Babo et al., 2018). Because this is a relatively recent field of research compared to traditional assessment, there have been fewer research studies looking into the perceptions of relevant parties towards online assessment compared to traditional assessment. Moreover, existing literature on the field is limited and has been focusing on the effects of student perceptions and perspectives toward online assessment rather than the teacher perceptions and perspectives on online assessment (Babo et al., 2018; Zhang et al., 2021). Thus, the concept of teacher perspectives of online assessment has been underresearched, especially in Turkey. In this section, relevant research on online assessment will be explained thematically under the title of studies on teacher perceptions regarding online assessment, studies on student perceptions of online assessment, and studies researching student and teacher perceptions of online assessment.

2.2.1. Studies on Teacher Perceptions Regarding Online Assessment

There are many noteworthy qualitative studies on teacher perceptions of online assessment. However, these studies are conducted with a limited number of teachers who may not represent the general population. On the other hand, there are also some quantitative studies on teachers' perceptions regarding online assessment, yet they fall short in number. Thus, the area is immature both in the number of studies available and in the scope of studies having been done. Besides, most available studies focus on general teacher perceptions or the benefits teachers see and the challenges they experience. Moreover, very few studies compare the results of participants in terms different variables such as age, gender or teaching experience.

The results of a limited number of studies indicate that one of the most common concerns for teachers regarding online assessment is the issue of security and academic

integrity. One study that deals with the teacher perspectives in terms of academic integrity is Rogers' (2006) study which aims to find if online assessment tools are used in a higher education institution and what concerns the faculty has in terms of online education. With a questionnaire, Rogers (2006) revealed that more than half of the faculty used online assessments under unsupervised environments, and almost half of the faculty members using online assessment were suspected of cheating in varying methods such as Internet surfing during an assessment, copying, or cheating from others. It was also found that no instructors were using security software to prevent cheating (Rogers, 2006). Another recent study by Sa'di et al. (2021) conducted in certain universities in Jordan revealed that instructors were skeptical about online assessment due to security and academic integrity issues and the lack of training and expertise. Through an online survey, participants expressed their perceptions of online assessment and provided feasible solutions to the challenges such as providing training for online assessment practices for instructors, using high-tech plagiarism software, and using a combination of formative and summative online assessment tasks (Sa'di et al., 2021). Another study focusing on academic integrity in online assessment is Mellar et al. (2018)'s study. In their study, they used a mixed method of surveys and interviews to see if faculty made use of a newly introduced security system to prevent cheating cases in three different universities and some solutions to address the issue. They found that faculty expected cheating to be greater in online assessment, but it was not. Also, the biggest cheating cases occurred in the form of ghostwriting, plagiarism, or copying work from the Internet. Thus, they concluded that online systems do not increase cheating cases, but authorship-checking software should be used and assessment should be made in a variety of methods, rather than in one method or one type (Mellar et al., 2018).

According to other studies investigating how teachers perceive online assessment in general, teachers tend to have positive attitudes regarding online assessment in general but also have certain concerns such as the lack of technical infrastructure, technical and technological support, or security. For instance, in a study by Chien et al. (2014), semi-structured interviews were conducted to explore teacher beliefs about technology-based assessments and the relation between teacher beliefs and their practices. It was found that the vast majority of teachers found technology-based assessment useful, beneficial, and

effective, and the difficulties regarding the use of technology-based assessment stemmed from poor infrastructure or lack of technical support (Chien et al., 2014). A more recent study by Küppers & Schroeder (2020) looked into university teachers' perceptions of online assessment through online surveys and demonstrated that most of the teachers were open-minded about the use of online assessment, and their major concerns were related to fairness and security. They also compared demographic results and revealed that the younger and the more technologically experienced the teachers were, the more positive they had toward using online assessment tools.

Other studies investigating teacher perceptions of online assessment show that although teachers find online assessment useful in times of crisis, they do not prefer to use it as a regular assessment. For instance, a study in Saudi Arabia explores the university instructors' general feelings toward online assessment in terms of its validity, reliability, security, practicality, the types of questions they prefer during online assessment, and differences between the types of questions that male and female instructors choose to use (Alghammas, 2020). In his quantitative study, Alghammas (2020) used Dermo's (2009) questionnaire, which originally explored student perceptions on the issue and found that instructors working at Saudi universities had a slightly positive attitude toward the use of online assessment at universities due to some concerns such as technical problems, security issues, and reliability. It was also revealed that the faculty had not used online assessment tools a lot previously. The research indicated that online assessment might be useful in difficult times but may not stand as a regular assessment method in their institution. The research could not indicate any significant correlation between the gender of participants and their question type preferences. As for the types of questions, most faculty members expressed that the questions should be feasible for objective grading due to immediate feedback opportunities and scoring ease (Alghammas, 2020). In another noteworthy and recent study, Yulianto & Mujtahid (2021) explored teacher perceptions towards online assessment through online interviews with 12 teachers and found out that in the Indonesian context, online assessment was less effective than traditional assessment due to the socio-economic background of students, lack of Internet and teachers' inexperience in and unfamiliarity with the technology. They found that the teachers found

online assessment useful in emergency situations but very difficult to conduct (Yulianto & Mujtahid, 2021).

Some other studies indicate that teachers might also have negative attitudes toward online assessment due to several factors such as inexperience in technology, lack of support from relevant parties, or personal barriers or difficulties. In a study that investigates teachers' perceptions of online assessment in higher education institutions in Lebanon, it was found through online interviews that instructors were anxious about using online assessment due to the lack of training before using it, and they were refraining from using summative assessment via online tools (Mirza, 2021). In China, another qualitative study by Zhang et al.(2021) revealed that the lack of preparation time and training caused a great variety in teacher practices of online language assessment at universities. The sudden change to online assessment due to unexpected situations caused great stress among teachers. Teachers also stated that they feared security issues, and thus, they mostly used formative online assessment rather than summative online assessment (Zhang et al., 2021). Another study in the Indonesian context searched for teacher perspectives of online formative assessment and the advantages and constraints of online assessment according to their understanding (Astiandani & Anam, 2021). Through semi-structured interviews, it was found that public school teachers mostly had negative perceptions toward online assessment due to the lack of parental support when necessary and the irresponsible behaviors of students. In private institutions, though, teachers were neutral toward it. Although they listed advantages such as immediate feedback, promoting autonomy, and being enjoyable and motivating for the students, the lack of Internet connection and the time-consuming nature of creating assessments made online assessments difficult for them. Teachers also proposed some solutions such as getting support from all stakeholders and obtaining better Internet connection availability (Astiandani & Anam, 2021). In Iran, a similar result was found in Ghanbari and Nowroozi's (2021) study which searched for teacher perceptions of online assessment through interviews and showed that teachers faced technological barriers such as lack of technical infrastructure and lack of technical knowledge and thus, their online assessment experience was affected negatively. Teachers also expressed personal problems such as lacking motivation and awareness about the issue. Thus, their overall perspective on online assessment practices remained negative

(Ghanbari & Nowroozi, 2021). Another similar result was obtained from a study conducted in India by Joshi and others (2020) through an interview with 19 teachers which found that teachers had problems in both home settings such as lack of basic facilities, personal external distractions, and in institutional settings such as lack of budget, training and technological support (Joshi et al., 2020).

There are also a few studies looking into gender differences in perceptions toward online assessment, and apparently, studies provide contradictory results in this aspect according to different countries. A quantitative study by Öz (2014b) in the Turkish context conducted to explore pre-service English teachers' perceptions of web-based assessment found that female students were less anxious using online assessment and were more likely to use it in their future studies. It was also shown that although most student-teachers had positive attitudes towards online assessment, but the majority of those were not likely to use it in their future practices. It was also revealed that the more computer-literate they were, the easier they found online assessments to use (Öz, 2014b). Another significant study (Abduh, 2021) which also investigated English teacher perceptions of online assessment in the Saudi Arabia context, found through a quantitative analysis of the data that the perceptions of English teachers of online assessment did not indicate a meaningful distinction regarding gender. It was also revealed that teachers had a moderate attitude toward online assessment due to the challenges such as technical problems, cheating issues, and limitation of productive skills assessment during the online assessment. (Abduh, 2021).

2.2.2. Studies on Student Perceptions Regarding Online Assessment

As mentioned earlier, the number of studies investigating student perceptions and attitudes toward online assessment is higher when compared to the number of studies investigating teacher perceptions. Moreover, they are varied in their method of data collection as there is a sufficient number of qualitative, quantitative, and mixed-method studies. The results of these studies indicate that there are contradictory results in terms of student perceptions of online assessment.

Results of a number of studies indicate that students enjoy being assessed online, and they find online assessment engaging in general. A study (Bernik & Jereb, 2006)

aims to find the readiness level of students for online assessment through online questionnaires conducted in years 2004 and 2005 twice with two different groups of students. The results suggest that according to the pre-test survey, students were ready for online assessment tasks and a post-test survey showed that students found online assessment quite interesting and appealing. It was also found that there was less possibility of cheating and not many problems with technology or with infrastructure which are in fact controversial with some other studies. As a result, it was indicated that online assessment is preferable for students compared to traditional assessment (Bernik & Jereb, 2006). Another study which indicated that students enjoy online assessment is Bailey and others' (2015) study. With an online survey, university students' perceptions toward online assessment methods were studied, and the results showed that the majority of students had a positive experience in terms of online assessment; they thought that it increased their levels of motivation and engagement, and they mostly found video activities, Twitter summaries and interviews enjoyable and engaging (Bailey et al., 2015). In another quantitative study (Alsadoon, 2017), it was seen that student perception of online assessment in Saudi Arabia was quite good, and students had positive attitudes toward online assessment, mainly because of its certain benefits such as immediate feedback and unbiased grading opportunities (Alsadoon, 2017). A more recent study by Mahbub (2020) indicated in a mixed-method study, combining a questionnaire and interviews that students positively appreciated assessment through online gamification. Twenty-one Indonesian university students expressed that online assessment through gamification contributed to their learning positively, and they felt quite comfortable during assessment (Mahbub, 2020). One other study that studied online gamification to assess vocabulary is Hadijah et al. (2020)'s study which aimed to find students' perspectives on using an interactive test for vocabulary testing. It was found through semi-structured interviews that students became more active and enthusiastic with online assessment through gamification (Hadijah, 2020). To sum up, it can be seen that many studies have revealed that students have positive attitudes toward online assessment.

Some studies indicate that students have positive perceptions of online assessment to varying degrees, yet they still have some concerns such as security and technical failures in mind related to the use of online assessment. One of the most prominent studies in the

literature to explore student perspectives on online assessment is Dermo (2009)'s study in which an online questionnaire evaluating online assessment in six dimensions which are validity, reliability, security, practicality, affective factors, and teaching and learning was shared with 130 students who had experienced online assessment before. The results indicate that students generally had moderately positive attitudes towards online assessment and expected that online assessment would become a part of their education. Online assessment was found to be a valid, reliable, and practical way of assessment and added value to teaching and learning. However, the study showed that students had security issues in mind related to online assessment as they had a low level of trust in the online systems in terms of keeping their grades safe (Dermo, 2009). Dermo (2009)'s questionnaire was used in another study in the Palestinian context with the very same objective, and the results of the questionnaire with 342 Palestinian students indicated that besides its benefits in terms of reliability of grading and efficiency in time and effort, online assessment reduces stress and anxiety of students (Shraim, 2019). On the other hand, students had some reserves regarding online assessment such as security and fairness issues and possible technical problems (Shraim, 2019). Another study that investigates students' doubts over online assessment is Riera et al. (2018)'s large-scale study which used a combination of a questionnaire shared with 1200 students and small focus group interviews. The study revealed that although students had positive opinions on online assessment in general, they had concerns in mind such as lack of security and system failure issues (Riera et al., 2018). A significant study by Şanlı (2003) aimed to explore students' perspectives on the general online assessment-related opinions, user interface of the online assessment system and the impacts of the system on teaching and learning indicated that students had positive attitudes regarding online assessment. However, they also believed that security of the system must be improved and the website interface should be changed to make the online assessment experience easier (Şanlı, 2003). Another study with a similar focus belongs to Babo et al. (2018) which investigated student perceptions of the use of multiple-choice questions in online assessment. To this end, an online survey including survey items on time, difficulty, and affective factors was shared with 84 university students, and it was found that although students had a positive attitude toward online assessment, they had concerns regarding time limits and possibilities of system failure or technical issues (Babo et al., 2018). An extended version

of the same study (Babo et al., 2020) whose data collection lasted over five years with 815 students used a similar online survey to understand students' opinions and perceptions of online assessment with the use of multiple choice questions. The study revealed that students appreciated the use of multiple-choice items in online assessment. However, they concluded that online assessment is useful for theoretical topics but unsuitable for practical ones (Babo et al., 2020). Another significant study with a similar result is conducted in the Turkish context with a questionnaire and interviews to investigate the level of academic success in traditional and online assessments and understand student perceptions of online assessment (Ilgaz & Afacan Adanır, 2019). It was found that there was no significant difference between the grades of students in online or traditional assessment, and although students found online exams reliable, efficient, and useful, they had certain concerns such as the possibility of technical problems and time limit issues (Ilgaz & Afacan Adanır, 2019). One study by James (2016) which investigated the effects of an online supervised and invigilated exam showed similar results to ones found by Ilgaz and Afacan Adanır (2019) as the trial of the procedure with a group of students and the semi-structured interviews afterwards revealed that although students found online assessment useful, cost-effective and flexible, they expressed that they had many technical problems and they did not like being observed by an invigilator (James, 2016). As can be seen from the studies above, online assessment may be regarded as a positive assessment method by students; however, it might bring about some concerns for them at the same time.

Despite the benefits and positive sides of online assessment expressed by students as can be seen in the results of the studies mentioned above, a few studies show that students face serious challenges during online assessment or have serious concerns related to it. One such study belongs to Khan and Khan (2019) which used interviews to investigate student perceptions of online assessment in the university context. The results revealed that besides not valuing online assessment and not comprehending the need for it, technological inexperience of students led to a negative experience with online assessment for them and created distrust in technology and infrastructure (Khan & Khan, 2019). The researchers believe that the acceptance of students might increase over time with better technological infrastructure and technological expertise. Another important

study with a similar result is Bahar (2014)'s study. The responses of 273 students studying at the tertiary level in Turkey indicated that when students study online, they have more positive attitudes toward it, but when they study face to face and have exams online, this creates tension for students (Bahar, 2014). Students have serious concerns regarding online assessment such as possible technical problems, connection loss, or health-related issues such as screen fatigue (Bahar, 2014). One study with a completely different focus indicated that online assessment might also cause serious issues in terms of academic integrity. King et al. (2009) conducted a study to understand the attitudes of tertiary level students toward the issue of cheating in online assessment. They found that a high majority of the students had the idea that it was a lot easier to cheat online compared to traditional assessment. Also, almost all students felt free to cheat in online assessment if there were no set test-taking rules and policies.

As can be seen, there are several studies checking online assessment in terms of the student perspective. According to the studies mentioned above, it can be concluded that online assessment has many positive aspects for students, such as being enjoyable and engaging, flexible and time efficient, and improving teaching and learning practices in class. However, it can also be concluded that students share some common concerns such as security or the system or safety of grades, technical issues and problems, and not being suitable for certain topics. Thus, checking both student views and teacher views on the issue might be helpful for research purposes to come to more meaningful conclusions.

2.2.3. Studies Investigating both Teacher and Student Perceptions Regarding Online Assessment

Since they are the most important stakeholders of assessment, it is important to study the perceptions of teachers and students together. Besides, when these components are studied together, it is logical to assume that the research is carried out in the same context for both students and teachers as in the studies of Korkmaz (2019), Asma (2021), Hichour (2022), Rollim and Isaias (2018) and Fitriyah and Jannah (2021). Thus, it can be concluded that the study becomes a meaningful and consistent one when teacher and student perspectives are studied together.

Although there is not ample research studying teacher and student perceptions of online assessment together, available literature indicates that both students and teachers

have positive perceptions of and certain concerns about online testing and assessment. In their study, Fitriyah and Jannah (2021) investigated teacher and student perceptions toward online assessment in the Indonesian EFL context through a quantitative study with students and a qualitative study with teachers. They found that both teachers and students had positive attitudes toward online assessment. According to the students, online assessment possessed advantages of flexibility, freedom of time and space, and learner autonomy. Students' concerns were more related to administration complexities, reduced interaction with their teachers, and test anxiety. As for the teachers, it was found that online assessment was flexible, time efficient, and full of new learning opportunities for students and teachers. The concerns they had about online assessment were the possibility of cheating, the difficulty of test preparation, lack of training, and possible connection problems for students (Fitriyah & Jannah, 2021). Another study also revealed that while both parties had positive perceptions of online assessment, they also held certain concerns. In his study, Fageeh (2015) investigated teachers' and students' attitudes toward online assessment practices in the Saudi Arabian context through a quantitative study with 400 students and 25 teachers. He found that students enjoyed the online assessment system and felt confident and relaxed with their experiences during assessment. However, they also had issues with the Internet connection, reliability of their grades, and the ease of cheating. Likewise, teachers enjoyed using online assessment, and believed that the system was helpful for teaching and learning purposes. Nonetheless, they had issues with logistics such as arranging a common time and venue with the students, student discipline during assessment, and connection problems (Fageeh, 2015). Another important study that investigates both students' and teachers' perceptions of online assessment belongs to Asma (2021). She followed a case-study method for her thesis to understand if teachers and students had positive attitudes toward online assessment with 134 students and 26 instructors who all studied and worked in the same institution. Her study revealed that both parties stated benefits and concerns regarding online assessment. Students stated that online assessment provided them with a less stressful atmosphere and flexibility of time and space. However, they also suffered from technical problems, ample cheating opportunities, concentration problems, and time limitations. As for teachers, they found the system to be flexible, eco-friendly, and time-saving. However, they also stated they experienced serious cheating issues and connection problems which made invigilating

difficult. Grading was also difficult, and less social interaction with students and with each other created stress for the teachers. Another negative point teachers stated was screen fatigue resulting from a long time of computer exposure (Asma, 2021). A similar finding can be seen in Hichour's (2022) study aiming to investigate Algerian teachers' and students' perspectives of online assessment experiences with Moodle, an online learning management system. In his mixed method study which investigated teachers' and students' perspectives with questionnaires and compared students' grades, it was found that both teachers and students had positive attitudes toward online assessment. Both parties found online assessment as a valid, reliable, and secure method of assessment. However, both parties had concerns related to Internet access. Moreover, teachers stated that the system was open to cheating. When student grades were compared with traditionally assessed components, it was seen that students got higher grades in online assessment, possibly due to the availability of lesson resources, as stated in the study (Hichour, 2022).

Other studies indicate slightly different results, suggesting that students may merely have positive perceptions of online assessment, whereas teachers might have positive perceptions and concerns in mind regarding the issue. For example, in a mixed method study, Korkmaz (2019) investigated teacher and student perspectives toward online assessment of oral skills in the Turkish higher education context and found that although students felt relaxed, confident, and positive in online speaking assessment, teachers had concerns regarding reliability and practicality. They experienced rater reliability issues during assessment and found online speaking assessment labor intensive. They also stated that the lack of necessary training made the experience more difficult. On the other hand, teachers also stated that online speaking assessment was time-efficient and beneficial for student learning (Korkmaz, 2019). Another study with a similar result belongs to Meccawy et al. (2021) who investigated online assessment practices during the time of crisis through open-ended questionnaires and found that although students had positive attitudes toward online assessment mainly because of less stress and immediate feedback opportunities, teachers thought that the system was open to cheating cases and was not suitable for their lessons. Despite enjoying autocorrection opportunities, more than half of the teachers stated that they would still use traditional assessment methods

(Meccawy et al., 2021). Another study with a similar result belongs to Rollim & Isaias (2018), who studied student and teacher perceptions of online assessment through a quantitative study in the Portuguese higher education context. It was found that most students enjoyed increased motivation, flexibility, and accessibility and trusted online assessment systems, whereas teachers distrusted the security systems of online assessment and experienced some technical problems. While stating that the system provided them with flexibility and accessibility, the teachers could not observe direct improvement in student learning, and their workload did not decrease as much as they expected (Rollim & Isaias, 2018).

There are also studies that suggest that formative assessment works better in online environments. For example, in their study, Gaytan and MacEwan (2007) aimed to find which types of online assessment tasks and tools are perceived as effective by students and faculty. Their study revealed that the tasks that students found the most effective were mainly formative assessment tasks such as self-assessment tasks, online discussions, and portfolio tasks, and similarly, the tasks that were found the most effective by the faculty were projects, portfolio tasks, self-assessment tasks, and peer assessment tasks. It was also suggested that the majority of both students and teachers had positive attitudes toward online assessment (Gaytan & Mc Ewan, 2007). Another study that also suggested that formative assessment works better with online assessment belongs to Baleni (2015) which was carried out in the South African context. He used Dermo's (2009) questionnaire which investigated student perceptions of online assessment, and it was found that online assessment was more beneficial and efficient when formative assessment tasks were used. It was also stated that teachers felt comfortable doing online assessment tasks, and students enjoyed the less stressful and flexible atmosphere of online assessment with easy and quick feedback opportunities.

2.2.4. Summary

The results of the studies indicate that although online assessment is perceived as a generally positive assessment tool which is a valid, reliable, and practical method by both teachers and students, both parties expressed certain concerns regarding the concept. According to the results, teachers have fundamental concerns and great benefits in mind

about online assessment. In addition, students have explicit issues and certain advantages with regard to online assessment.

Teachers seem to have four main concerns related to online assessment; exam security, lack of expertise in online assessment, reliability concerns, and a shortage of essential IT support from relevant people. The issue of academic integrity seems to be the most serious concern that teachers have in mind regarding online assessment. In many cases and contexts, teachers suffer from a lack of basic software to prevent plagiarism and cheating. Thus, they express that they suspect multiple occurrences of cheating in various methods in their contexts. Moreover, as teachers have not used online assessment methods very often previously, their lack of expertise both in computer technologies and online assessment leads to serious issues for them. Therefore, in many institutions, teachers express that they need relevant support and training to be able to continue online assessment safely and effectively. For this reason, many teachers in many institutions only use formative assessment tasks in online contexts, which creates a reliability problem. Because not all teachers in institutions assess their students using a variety of tasks, the reliability of assessment decreases. Teachers also express that a lack of technical support and infrastructure such as reliable Internet connection and available Internet quota causes problems during assessment. Limited available access to the Internet makes online assessment extremely difficult for teachers. Besides, the lack of necessary technical support increases the level of challenge and decreases the teachers' motivation to conduct online assessment procedures.

Teachers also express that they are generally positive toward online assessment and state positive sides of online assessment mainly regarding practicality benefits. The first practicality benefit of online assessment is the scoring ease. When the assessment items are chosen to be objectively graded, scoring becomes easy and error-free. This provides the students with immediate feedback opportunities, which is perceived as a benefit by teachers because of increasing student motivation. Moreover, teachers find online assessment flexible as it is unbound to time-related and spatial concerns. By being free of time and space, online assessment becomes time-saving for teachers. Another practicality benefit of online assessment is expressed as being eco-friendly by reducing photocopying costs and other resources. Prevention of paper waste and photocopying

costs are greatly appreciated by the teachers. Overall, teachers express that online assessment is useful in times of crisis since it is the best solution available at hand.

Studies indicate that students share similar positive perceptions of online assessment related to practicality. They mention that online assessment offers them flexibility by being free of time and space. They also consider the reliability of online assessment tools as a positive side of online assessment. Immediate feedback opportunities increase their motivation and provide grading accuracy. According to students, online assessment also increases reliability by providing objective grading, thus, yielding to unbiased grading. Students mainly express online assessment as a less stressful method of assessment thanks to the interactive nature of tasks. Another benefit stated by students is lowered affective filters for them thanks to online assessment. They state that online assessment is enjoyable and motivating for them as the use of videos, gamification, discussion tasks, and other interactive formative assessment tools make them engaged and enthusiastic.

Despite perceived benefits, students have certain concerns about online assessment, which are mainly related to security, IT and health-related issues. A substantial number of students express that their main concern about online assessment is possible technical failures such as the lack of Internet connection, connection loss, or computer breakdowns during assessment. They state that these kinds of possible problems increase their level of anxiety and create tension for them. Another issue mildly related to technical problems is security, as students fear that their grades or answers to questions in the exam are safely kept and stored in the online assessment systems. They think that because of the lack of security, they have a low level of trust in online assessment systems. They state that a more secure online assessment system is necessary. Moreover, like teachers, students also believe that online systems create an exam athosphere that is more prone to cheating, which acts as a temptation for students to cheat. Another issue is related to their inexperience with the use of the Internet, online assessment systems, or the computer, and they state that this increases their stress and anxiety levels by leading to time limitation concerns. Moreover, being in front of the screen for a long time cause screen fatigue for students. In sum, although they seem to enjoy online assessment, students might have serious concerns related to it.

3. METHODOLOGY

This section discusses the methodology of the study in five subsections: research context, participants, data collection tool, procedures and data analysis. In each part, specific details regarding the research is explained in detail. A short overview of the study is explained in the research context, and the information regarding the participants is given in the participants subsection. The data collection tool describes the questionnaire used for the study and the procedures part explains how the study is carried out. Finally, the data analysis subsection provides information on how the data is analyzed.

3.1. Research Context

This research aims to identify online assessment perspectives of teachers working at English preparatory programs of universities in Turkey. It mainly explores instructors' general perspectives of online assessment. Following that, instructors' perspectives are analyzed against a number of variables such as age, gender, graduation department, educational background, years of experience in teaching, the type of institution they work at, their position at their workplaces, and their self-reported computer expertise to identify the relationships between them. The research follows an analytic approach as it is assumed that these variables relate to the predetermined construct of online assessment perspectives, and it tries to discover the relationship between this construct and the variables rather than attempting to analyze the construct as a whole (Dörnyei, 2003; Seliger & Shohamy, 1989). In nature, the research is descriptive as it attempts to explore and explain naturally occurring phenomena as in teachers' perspectives by looking at the relationships of different variables one by one, without any experimental design scheme or manipulation while using the scale. The participating groups naturally exist without any pre-formation; thus, the data is collected naturally (Seliger & Shohamy, 1989). In terms of the degree of control over the research context, it can be said that the research is carried out in a semi-controlled environment as the context is narrowed down to university teachers only, and the scope is to teacher perceptions of online assessment in general, as well as, reliability, security, practicality and pedagogy issues (Seliger & Shohamy, 1989). Finally, it is possible to claim that the researcher has been as objective as possible, since she had no control over the participants and their answers (Seliger & Shohamy, 1989).

Student Perceptions of e-Assessment Questionnaire (SPEAQ), originally designed by Dermo (2009) to investigate student perceptions toward online assessment is used to measure teachers' perspectives. Online questionnaires are useful and efficient methods of data collection as they are fast and easy to administer, they ignore geographical concerns, have financial advantages and provide anonymous data (Wright, 2005). Moreover, questionnaires are effective tools for gaining data on people's opinions, attitudes, beliefs, and values on concepts (Dörnyei, 2003).

3.2. Participants

Participants of the study were 302 English instructors working at the English preparatory programs in various universities in Turkey; 228 (75.5%) were females, and 74 (24.5%) were males. The mean age of these participants was 41.1, between 24 and 71. The mean score for teaching experience was 17.5 years with one year of experience being the lowest and 48 years being the highest level of experience. A hundred seventy seven of the participants had a master's degree (58.6 %), while 92 had a bachelor's (30.5%) and 33 had a doctoral degree (10.9%). Of these participants, one hundred eighty eight instructors graduated from English Language Teaching departments (62.3%), 74 of them graduated from English Language and Literature departments (24.7%), 19 from American Culture and Literature departments (6%), and 21 (7%) from other departments such as Translation Studies, or Linguistics. Of the participants, two hundred and three (67.4%) instructors worked at private or foundation universities, and 99 of them worked at state universities (32.6%). Two hundred fifty-six of these teachers expressed that they did not have an administrative duty (84.7%), while only 46 of them (15.2%) stated that they had administrative duties. As for office duties such as being a curriculum development, testing and assessment, or professional development unit member, two hundred and six instructors stated that they did not have such responsibilities (68.2%), and 96 of them stated that they were working at one of these offices (31.7%). One hundred sixty-eight of the participants stated that they found themselves good in terms of computer expertise (56.6%), 69 of them (22.8%) stated they were excellent at using computers, and 65 of them (21.5%) thought that they were adequate users of computers. More detailed information about the participants can be seen in *Table 6* below.

Table 6. Background Information of Participants

	Number	%	Mean	Standard Deviation
Gender				
Female	228	75.5		
Male	74	24.5		
Age				
20-30	36	11.9		
31-40	137	45.4		
41-50	77	25.5	41.1	.91
50+	52	17.2		
Teaching experience in years				
0-10	70	23.2		
11-20	129	42.7		
21-30	74	24.5	17.5	.90
30+	29	9.6		
Graduation department				
English Language Teaching	188	62.3		
English Language and Literature	74	24.5		
American Culture and Literature	19	6.3		
Other	21	7		
Highest degree received				
BA	92	30.5		
MA	177	58.6		
Ph.D.	33	10.9		
Type of institution				
State	99	32.8		
Private/ Foundation	203	67.2		
Administrative duty				
Yes	46	15.2		
No	256	84.8		
Office duty				
Yes	96	31.8		
No	206	68.2		

Computer expertise		
Adequate	65	21.5
Good	168	55.6
Excellent	69	22.8

3.3. Research Tool

The study used two data collection tools (see Appendix A). First, a background questionnaire to collect demographic and background information about participants was shared with the participants. The participants were expected to give information about their gender, age, the highest level of educational degree completed, graduation department, level of teaching experience in years, position in their institutions, and their level of computer expertise. The second tool was the Student Perceptions of e-Assessment Questionnaire (SPEAQ) developed by Dermo (2009), which was originally administered to students to identify their online assessment perceptions and perspectives. In the original research, the questionnaire was divided into six dimensions related to online assessment: affective factors, validity, practicality, reliability, security, and effects on learning to analyze the data more effectively (Bryman & Cramer, 2001 as cited in Dermo, 2009). Then, five indicators to measure students' perceptions of e-assessment for each dimension were formed in accordance with existing literature and expert opinions (Dermo, 2009). Although the overall reliability coefficient and construct validity values for the scale were not reported, the reliability coefficients in Cronbach's alpha for each component of the questionnaire were stated in the paper. Reliability value for affective factors was .80, and .33 for validity. For practicality, it was measured as .68, and for reliability, it was .63. For security, it was measured as .69, and for effects on learning, it was .82. Dermo's (2009) SPEAQ was adapted to measure teachers' perceptions and perspectives of online assessment and includes 30 items, slightly changed in wording from the original to fit the purpose of this research. To indicate their beliefs on each of the 30 statements with a numerical expression, the participants were asked to select the options given in a Likert-type scale of 5 points (5= "Strongly Agree"; 4= "Agree", 3= "Neutral", 2= "Disagree", 1= "Strongly Disagree"). Each of five statements of the survey refers to an aspect of testing and assessment. The first five statements (items 1,2,3,4,5- See Appendix A) collected data on affective factors regarding the use of online assessment. The next five

(items 6,7,8,9,10) investigated validity issues. Practicality was measured with items 11, 12, 13, 14, and 15. The next five items (items 16, 17, 18, 19, 20) measured reliability, and items 21, 22, 23, 24, and 25 collected data on security perceptions, and finally, items 26-30 measured the effects of online assessment on teaching and learning. Some of the statements (1, 3, 4, 7, 8, 10, 12, 13, 14, 17, 18, 19, 20, 23, 24, and 29) were formed negatively, and others (2, 5, 6, 9, 11, 15, 16, 21, 22, 25, 26, 27, 28 and 30) were formed positively, which affects the way the Likert scale was used for the study. According to Dermo (2009), the statements under each category would contribute to a reliable measure of the main aspect of testing and assessment as a whole. However, the main aspects or descriptors for the statements are not provided to the participants on purpose, as not to misguide them or affect their opinion (Dermo, 2009). Adapted versions of Dermo's (2009) SPEAQ scale have been used in several research studies worldwide to explore students' and teachers' perceptions of online assessment. The reliability coefficients calculated by Dermo (2009) and all the other researchers that used adapted versions of Dermo's (2009) SPEAQ scale can be found below in *Table 7*.

Table 7. Reliability Values of the Scale in Other Studies

	Studies	**Participants**	**Reliability Coefficients in Cronbach's Alpha**	
30 items	Dermo (2009)	Students	Affective Factors	.80
			Validity	.33
			Practicality	.68
			Reliability	.63
			Security	.69
			Effects on Learning	.82
			Overall Reliability	Not reported
18 items	Shraim (2019)	Students	.70	
17 items	Ilgaz & Afacan Adanır (2019)	Students	.92	

30 items	Alghammas (2020)	Teachers	Not reported
15 items	Baleni (2009)	Teachers and Students	Not reported

3.4. Procedure

Upon receiving the Educational Sciences Ethics Committee approval at İstanbul Medeniyet University (see Appendix B), the online survey which consists of two parts (Part 1: Demographic Questions; Part 2: Scale Questions), was shared with instructors working at English preparatory programs of diverse universities in Turkey via e-mails and social media tools. Since the online questionnaire and scale are one of the most efficient ways of data collection, participation is positively affected when participants are sent personal messages via mail (Dermo, 2009; Muñoz-Leiva et al., 2010). The data were collected through personalized e-mails.

The opening page of the scale, the Teacher Perceptions of Online Assessment Questionnaire consisted of a brief overview of the study, informing participants about the aim of the survey, the time frame required for the completion of the survey and contact information of the researcher. The same page also ensured that the data would be anonymous, used only for research purposes, and participation is on a voluntary basis. The items to collect demographic information were presented on the next page and the survey questions followed the demographic information page. Before starting the questionnaire, the participants knew that they could leave the survey anytime they wanted without submission. All the questions in the questionnaire were compulsory to answer, and if participants left the survey without completing items, they could come back anytime they wanted to start back from where they had left, which provided the participants with the flexibility of time and space. Having collected the data, which took around one month, the questionnaire was deactivated, and participants could not take the questionnaire from that point on.

3.5 Data Analysis

In the analysis process, Statistical Package for Social Sciences (SPSS) data analysis software was used. The data were analyzed for nominal variables such as gender division, graduation department, workplace, highest degree received, institutional responsibilities and self-reported computer expertise, and for ordinal variables which were age and teaching experience. For ordinal variables, which were age and teaching experience, mean, minimum and maximum values, and standard deviation were calculated. Right after that, intervals for age and years of teaching experience were specified. For variables of age, highest educational degree, department of graduation, workplace, institutional responsibilities, and self-perceived computer expertise, frequencies and percentages were computed as well.

Before the analysis, some of the items that have negative expressions regarding different aspects of online testing and assessment (items 1, 3, 4, 5, 7, 8, 10, 12, 13, 14, 17, 18, 19, 20, 23, 24, 28 and 29) were reversed [1=strongly disagree] became 5 [strongly agree], 2 [disagree] became 4 [agree], 3 [neutral] stayed the same, 4 [agree] became 2 [disagree], and 5 [strongly agree] became 1 [strongly disagree]), so that disagreeing with items not favoring online testing and assessment would be similar to agreeing with the positive aspects of the concept. The reliability of the overall survey with 30 items was found as $\alpha = .92$, indicating good internal reliability. The reliability value of each aspect of online assessment is as follows: $\alpha = .81$ for affective factors; $\alpha = .61$ for validity; $\alpha = .73$ for practicality; $\alpha = .70$ for reliability; $\alpha = .73$ for security and $\alpha = .83$ for impact on teaching and learning. The overall construct validity of the scale was computed as a % of total variance of 59.82%. The construct validity values for the aspects related to online assessment evaluated in this scale are as follows: 57.82% for affective factors; 62.84% for validity; 49.83% for practicality; 56.23% for reliability; 59.96% for security, and 60.31% for effects on teaching and learning. In order to compute the general perception score of the participants, their average scores for 30 items were calculated, and participants with an average of 3.34-5 were found to have a positive attitude towards online testing and assessment. An average score of 1.67- 3.33 indicated a neutral perspective, and the mean score of 0-1.66 referred to a negative perspective of online testing and assessment. Afterwards, the frequency, mean and percentage of participants, and standard deviation

values for each item were computed separately. As most of the data had a normal distribution, the relationships between the survey items and all the variables except for the department of graduation were analyzed and investigated through parametric tests of One-way ANOVA and independent samples T-tests. The non-parametric variable of graduation department was analyzed with Kruskal-Wallis H tests. The results which were accepted to be statistically significant had a significance value smaller than .05 (p<0.05).

Table 8. The reliability coefficients and % of variances

Scales	N of Items	Cronbach's Alpha	% of Variance
Overall Scale	30	.92	59.82
Affective Factors	5	.81	67.82
Validity	5	.61	62.84
Practicality	5	.73	49.83
Reliability	5	.70	56.23
Security	5	.73	59.96
Impact on Teaching and Learning	5	.83	60.31

4. RESULTS

This section explains the research findings after careful analysis of the statistical data. In order to discover the answer to the first question, which aims to provide insight into instructors' overall perspectives of online assessment, the mean scores of participants were calculated, and the participants were put into three categories: having a positive, neutral, or negative perspective toward online assessment. As the second research question aims to identify relationships with and differences between instructors' perspectives toward online assessment and certain variables, parametric and non-parametric tests were conducted. The findings are presented separately under the research questions below.

4.1. RQ 1. How do EFL instructors perceive using online testing and assessment?

The values in *Table 9* show that instructors had an overall neutral perspective of online assessment in this context. According to the table, it can be seen that the impact of online assessment on the teaching and learning category had the highest average score ($\bar{x}$=3.38), whereas the security aspect had the lowest ($\bar{x}$=2.32). While the overall average for all the items was 2.81, the closest mean score to the overall average belonged to validity ($\bar{x}$=2.80) and practicality ($\bar{x}$=2.72).

Table 9. Descriptive statistics for the aspects of online assessment (N=302)

Scales	Mean	Std. Deviation
Overall Scale	2.81	.58
Affective Factors	2.70	.82
Validity	2.80	.69
Practicality	2.72	.74
Reliability	2.96	.72
Security	2.32	.67
Impact on Teaching and Learning	3.38	.70

According to *Table 9* and values given in Appendix C, the mean scores indicate that the perceptions of instructors could be different according to each component of online assessment. To begin with, the mean score of affective factors ($\bar{x}$=2.70) was lower than the overall average ($\bar{x}$=2.81), suggesting that teachers had psychological barriers

regarding online assessment. Instructors felt more comfortable with paper-based exams ($\bar{x}$=3.80) when compared to online exams ($\bar{x}$=2.36) despite the fact that they expected online assessment to be a part of the regular assessment at the tertiary level ($\bar{x}$=3.40), which is indicated in Appendix C. For the validity aspect of online assessment ($\bar{x}$=2.80), *Table 9* illustrates that instructors' perspectives were in line with their overall perceptions ($\bar{x}$=2.81). However, they possibly thought that online assessment could not effectively assess their subject area ($\bar{x}$=2.92), as English could be too complex to be dealt with online multiple-choice items ($\bar{x}$=3.21) and online assessment also tested IT skills of students ($\bar{x}$=3.40). Practicality ($\bar{x}$=2.72) also fell behind their overall perceptions of online assessment ($\bar{x}$=2.81) as they very strongly believed that technical problems ($\bar{x}$=3.96) and Zoom/computer fatigue ($\bar{x}$=3.59) made online assessment impractical. They did not seem to appreciate the practicality of online assessment in terms of time and space very much as well ($\bar{x}$=2.95). Yet, they welcomed the prevention of paper-waste with online assessment ($\bar{x}$=3.74). Reliability scores ($\bar{x}$=2.96) of instructors were a little higher than their overall perceptions ($\bar{x}$=2.81) as they thought that computer-based marking was more accurate ($\bar{x}$=3.50). However, they also thought that paper-based exams were fairer than online assessments ($\bar{x}$=3.45). Security ($\bar{x}$=2.32) fell significantly behind the overall perceptions of instructors ($\bar{x}$=2.81), making this aspect the most negatively perceived aspect of online assessment. They especially worried about the ease of cheating ($\bar{x}$=4.26), and they had little trust in the system in terms of plagiarism and cheating ($\bar{x}$=2.26) and hackers ($\bar{x}$=3.75). Instructors seemed to value the impacts of online assessment on teaching and learning as the mean score of this component ($\bar{x}$=3.38) was a lot higher than their overall perception ($\bar{x}$=2.81). Also, they seemed to appreciate the immediate feedback opportunity of online assessment ($\bar{x}$=3.65) and its complying nature with online learning ($\bar{x}$=3.75).

4.2. RQ2. Is there a significant difference between the perceptions of EFL instructors toward online assessment with regard to certain variables?

4.2.1. Gender

As *Table 10* indicates, gender is a significant differentiating variable for instructors' perceptions of online assessment (p=.00). It can be observed from their mean differences that males had a more positive attitude ($\bar{x}$=2.99) toward online assessment

compared to females ($\bar{x}$=2.76). Gender was a significant factor not only in the overall perspectives of instructors but also in the sub-components of online assessment, except for validity, as can be seen in *Table 10 (*p=.00-.01).

Table 10. Relationship Between Online Assessment Perceptions and Gender (Independent Samples T-Test)

	Gender	**N**	**Mean**	**Std. Deviation**	**F**	**Sig. (p-value)**
Online Assessment Perception	Female	228	2.76*	.55	7.98	.00
	Male	74	2.99*	.65		
Affective Factors	Female	228	2.62*	.77	5.53	.00
	Male	74	2.92*	.93		
Validity	Female	228	2.74	.68	.12	.72
	Male	74	2.93	.72		
Practicality	Female	228	2.66*	.71	1.92	.01
	Male	74	2.90*	.78		
Reliability	Female	228	2.90*	.68	4.78	.01
	Male	74	3.14*	.81		
Security	Female	228	2.27*	.64	5.60	.01
	Male	74	2.48*	.76		
Impacts on Teaching and Learning	Female	228	3.32*	.68	.01	.00
	Male	74	3.58*	.72		

As *Table 10* indicates, gender is a significant factor regarding the differences between the online assessment perceptions of instructors with a *p*-value of .00 in terms of affective factors. Similar to their results of overall online assessment perceptions, males had a more positive perception of online assessment ($\bar{x}$=2.92) than females ($\bar{x}$=2.62) regarding their emotions such as comfort, anxiety, or expectations. When individual items of affective factors are checked, *Table 11* indicates that differences were obvious in items

1 (*p=.00*), 2 *(p=.00*), and 5 (*p=.02*). Apparently, the use of computers created more stress for females ($\bar{x}$=3.33) than for males ($\bar{x}$= 2.85) according to item 1 shown in *Table 11*. Moreover, item 2 indicates that males ($\bar{x}$=3.66) expected computers to be used as regular assessment tools at a university more than females did ($\bar{x}$=3.32), and they ($\bar{x}$=2.60) felt more comfortable while doing exams on computers than females did ($\bar{x}$=2.28), as item 5 suggests.

As evident in *Table 10*, instructors' perceptions of the practicality aspect of online assessment differ significantly according to their gender (*p=.01*). Males ($\bar{x}$=2.90) had a more positive perspective of practicality than females ($\bar{x}$=2.66). This was also reflected in their responses to scale items, as *Table 11* indicates. Males ($\bar{x}$=3.72) did not believe that technical problems made online assessment impractical as much as females ($\bar{x}$=4.03) did. Also, computer or Zoom fatigue did not seem to affect males ($\bar{x}$=3.28) as much as females ($\bar{x}$=3.70), and they ($\bar{x}$=3.22) appreciated the spatial and time-related flexibility that online assessment creates more than females did ($\bar{x}$=2.86).

The analysis also indicates a significant difference in the mean scores of male and female instructors in reliability with a *p*-value of .01, as apparent in *Table 10*. Similar to practicality, males ($\bar{x}$=3.14) perceived online assessment as a reliable method of assessment more than females did ($\bar{x}$=2.90). Also, the analysis of scale items evaluating reliability in *Table 11* indicates significant differences. The mean differences between the genders show that males ($\bar{x}$=3.81) trusted the online assessment marking reliability more than females did ($\bar{x}$=3.41). Similarly, males ($\bar{x}$=2.91) believed that online assessment technology was reliable, while females held a slightly negative perception towards it ($\bar{x}$=2.58).

As *Table 10* illustrates, the difference between male and female instructors' perception of security in online assessment is significantly different as well (*p*=.01). The results suggest that although both genders' scores were lower than the mean score of their overall perceptions ($\bar{x}$=2.81), males ($\bar{x}$=2.48) seemed to have a more positive attitude toward security in online assessment compared to females ($\bar{x}$= 2.27). The analysis of the scale items showed that item 24 presented a significant difference (*p*=.00), suggesting that

females ($\bar{x}$= 3.85) perceived online assessment as more open to threats from hackers than males did ($\bar{x}$=3.41).

As for the impacts of online assessment on teaching and learning, it can be seen that there is a significant difference between the genders (p=.00), as shown in *Table 10*. As can be seen in *Table 11*, females ($\bar{x}$= 3.03) did not believe that online assessment could add value to learning as much as males did ($\bar{x}$= 3.40), and unlike males($\bar{x}$= 2.82), they thought that online assessment did not benefit learning and teaching much ($\bar{x}$=2.43). Similar to the other results, males ($\bar{x}$= 3.58) seemed to believe that online assessment impacted teaching and learning more positively than females ($\bar{x}$= 3.38). Scale analysis of this category also indicated significant differences for items 28 and 29.

Table 11. Independent Samples T-Test Analysis for Items Indicating Statistically Significant Difference in Gender

	Items	Gender	N	Mean	Std. Deviation	F	Sig. (p-value)
Affective Factors	1. Using a computer adds to the stress of exams for teachers.	Female	228	3.33*	1.03	9.58	.00
		Male	74	2.85*	1.27		
	2. I expect computers to be used as part of regular assessment at university.	Female	228	3.32*	1.04	.31	.00
		Male	74	3.66*	1.11		
	5. I'd rather do exams on a computer than on paper, because I am used to working online.	Female	228	2.28*	.96	12.09	.02
		Male	74	2.60*	1.23		
Practicality	12. Technical problems make online exams impractical.	Female	228	4.03*	.88	12.93	.01
		Male	74	3.72*	1.11		
	13.Computer/Zoom/Internet fatigue makes online assessments impractical.	Female	228	3.70*	1.04	.96	.00
		Male	74	3.28*	1.07		

	15. Online exams are more practical than paper based exams because they are free from time and space.	Female	228	2.86*	1.12	2.31	.01
		Male	74	3.22*	1.23		
Reliability	16. Marking is more accurate, because computers don't suffer from human error.	Female	228	3.41*	1.00	5.49	.00
		Male	74	3.81*	.93		
	17. The technology used in online assessments is unreliable.	Female	228	2.91*	1.04	2.17	.02
		Male	74	2.58*	1.14		
Security	24. The online exam system is vulnerable to hackers.	Female	228	3.85*	.86	7.46	.00
		Male	74	3.41*	1.03		
Impacts on Teaching and Learning	28. Online assessment can add value to students' language learning.	Female	228	3.03*	.90	1.20	.03
		Male	74	3.40*	.93		
	29. Online assessment is just a gimmick that does not really benefit learning and teaching.	Female	228	2.82*	.96	.20	.03
		Male	74	2.43*	1		

4.2.2. Age

According to *Table 12*, age is an effective variable that indicates significant differences among instructor perceptions of online assessment (p=.05). There was a significant difference in the perceptions of online assessment between the instructors aged 31-40 and the ones above 50. Apparently, instructors above 50 years old had more negative perceptions ($\bar{x}$= 2.63) of online assessment compared to the ones in the 31-40 group ($\bar{x}$=2.89).

For further insight, the relationship among age groups and their perceptions of affective factors, reliability, validity, security, practicality, and impacts on teaching and

learning in online assessment were analyzed. As can be seen in *Table 12*, the aspects indicating no significant difference were reliability (p=.52), validity (p=.26), and security (p=.55). On the other hand, the components that indicate significant differences are affective factors (p=.04), practicality (p=.05), and the impacts of online assessment on teaching and learning (p=.00). The instructor perceptions of affective factors in online assessment showed significant differences with regard to age (p=.05) as can be seen in *Table 12*. According to the results, younger teachers who were between 20-30 years of age ($\bar{x}$=2.83) had fewer affective barriers compared to those who were older than 50 years of age ($\bar{x}$=2.45), meaning they had more positive perceptions toward online assessment. A similar significant difference is obvious in *Table 12* on practicality and age (p=.05). The mean differences suggest that younger instructors (20-30:$\bar{x}$=2.74; 31-40:$\bar{x}$=2.79; 41-50:$\bar{x}$=2.74) found online assessment more practical compared to instructors older than 50 years of age ($\bar{x}$=2.46). *Table 12* also indicates that another aspect that shows a significant difference between instructors with regard to age was the impacts of online assessment on teaching and learning (p=.00). As can be seen, instructors between 31-40 years old ($\bar{x}$= 3.58) tended to have a more positive perception in terms of its benefits in teaching and learning compared to those older than 50 years old ($\bar{x}$= 3.05).

Table 12. Relationship Between Online Assessment Perceptions and Age (One-way ANOVA)

	Age Groups	**N**	**Mean**	**Std. Deviation**	**F**	**Sig. (p-value)**
Online Assessment Perception	20-30	36	2.86*	.60	2.62	.05
	31-40	137	2.89*	.56		
	41-50	77	2.79*	.65		
	50+	52	2.63*	.48		
Affective Factors	20-30	36	2.83*	.99	2.65	.04
	31-40	137	2.79*	.75		
	41-50	77	2.64	.85		

	50+	52	2.45*	.78		
Practicality	20-30	36	2.74	.67		
	31-40	137	2.79*	.74	2.59	.05
	41-50	77	2.74	.80		
	50+	52	2.46*	.62		
Impact on Teaching and Learning	20-30	36	3.33	.67		
	31-40	137	3.51*	.67		
	41-50	77	3.38*	.74	5.50	.00
	50+	52	3.05*	.64		
Validity	20-30	36	2.91	.13		
	31-40	137	2.86	.05	1.33	.26
	41-50	77	2.75	.08		
	50+	52	2.68	.08		
Reliability	20-30	36	3.03	.66		
	31-40	137	2.98	.74		
	41-50	77	2.98	.75	.74	.52
	50+	52	2.83	.65		
Security	20-30	36	2.25	.64		
	31-40	137	2.38	.66		
	41-50	77	2.27	.75	.70	.55
	50+	52	2.28	.64		

Values in *Appendix D* indicate further significant differences among instructors of different age groups in some items of the scale. Firstly, items 4 (p=.05) and 5 (p=.05) of affective factors showed significant differences. It seems that instructors who were older than 50 found doing exam-related tasks more difficult than younger instructors who were

between 20-30 years of age, and they preferred online assessment less compared to instructors of 31-40 years of age. Although no significant difference was found for overall validity perceptions of instructors (p=.26), further analysis of validity items revealed that there was a significant difference between the age groups of 20-30, 31-40, and 50+ for the item number 6 (p=.00). As can be seen in *Appendix D* instructors who were between 20-30 ($\bar{x}$= 3.11) and 31-40 ($\bar{x}$= 3.09) years of age thought that online assessment was more appropriate for their subject area compared to the ones older than 50 ($\bar{x}$=2.51). Moreover, item 11 in the practicality component indicated a significant difference between instructors of 20-30 and 31-40 years of age and instructors older than 50. It can be thought that instructors older than 50 ($\bar{x}$=3.32) were less sensitive in the use of paper in exams than instructors who were between 20-30 ($\bar{x}$=3.88) and 31-40 years old ($\bar{x}$=3.91). The same table also indicates that items 26 (p=.00) and 28 (p=.01) in the impacts of online assessment on teaching and learning components indicated differences. It is evident in the table in *Appendix D* that instructors older than 50 years old ($\bar{x}$= 3.28) did not believe that the immediate feedback opportunity of online assessment benefitted student learning as much as instructors between 20-30 years of age ($\bar{x}$=3.83) and 31-40 years of age ($\bar{x}$= 3.76) did. Similarly, instructors older than 50 ($\bar{x}$= 2.78) did not appreciate the potential value that online assessment could add to learning as much as instructorse between 31-40 years of age ($\bar{x}$= 3.28) did.

4.2.3. Teaching Experience

Another significant discriminator for instructors' online assessment perceptions is teaching experience in years with a significance value of p=.03, as *Table 13* indicates. According to the table, mean scores indicate that instructors with less experience in years had a more positive perception of online assessment than instructors with longer years of experience. As *Table 13* indicates, the most significant difference can be observed in the mean scores of instructors with less than 10 years ($\bar{x}$=2.93) and more than 30 years of ($\bar{x}$=2.57) experience, the latter being less positive towards online assessment compared to the former.

Table 13. Relationship Between Online Assessment Perceptions and Teaching Experience in Years (One-way ANOVA)

	Experience in Years	N	Mean	Std. Deviation	F	Sig. (p-value)
Online Assessment Perception	<10	70	2.93*	.60	2.84	.03
	11-20	129	2.83*	.55		
	21-30	74	2.77	.66		
	>30	29	2.57*	.38		
Affective Factors	<10	70	2.89*	.85	5.66	.00
	11-20	129	2.73*	.74		
	21-30	74	2.66*	.91		
	>30	29	2.17*	.64		
Validity	<10	70	2.92	.75	1.06	.36
	11-20	129	2.81	.67		
	21-30	74	2.74	.73		
	>30	29	2.69	.48		
Practicality	<10	70	2.80*	.77	2.78	.04
	11-20	129	2.77*	.70		
	21-30	74	2.68	.80		
	>30	29	2.37*	.355		
Reliability	<10	70	3.13	.73	1.83	.14
	11-20	129	2.92	.71		
	21-30	74	2.92	.74		
	>30	29	2.84	.62		
Security	<10	70	2.32	.66		
	11-20	129	2.33	.66		

	21-30	74	2.31	.78	.04	.98
	>30	29	2.28	.51		
Impacts on Teaching and Learning	<10	70	3.53*	.72	3.32	.02
	11-20	129	3.42*	.67		
	21-30	74	3.31	.75		
	>30	29	3.07*	.53		

Table 13 also points out that the variables of validity, reliability, and security reveal no significant differences among instructors with different levels of experience, as the *p*-values were .36, .14, and .98, respectively. This means that instructors with different years of experience had similar perceptions of online assessment in terms of validity, reliability, and security aspects. However, instructor perceptions of affective factors, practicality, and impacts of online assessment on teaching and learning indicated significant differences among different experience groups.

According to *Table 14*, affective factors indicate significant differences among groups according to their teaching experience. A post hoc analysis of the data revealed that instructors with less than ten years of experience ($\bar{x}$= 2.89) had a more positive perception of online assessment than those with more than 30 years of experience ($\bar{x}$= 2.17). A similar difference can be observed in instructors with 11-20 years of experience ($\bar{x}$= 2.73) and 21-30 years of experience ($\bar{x}$= 2.66) and the ones with more than 30 years of experience ($\bar{x}$= 2.17). The results indicated that instructors with less than 30 years of experience had lower affective barriers compared to the ones with more than 30 years of experience.

Table 14 further reveals that teaching experience plays an important part in instructors' practicality perceptions of online assessment since a statistically significant difference was obtained with analysis (p=.04). The mean scores of different teaching experience groups indicated that instructors with less than ten years of experience ($\bar{x}$= 2,80) and between 11-20 years of experience ($\bar{x}$= 2.77) found online assessment more

practical than instructors with more than 30 years of experience ($\bar{x}$= 2.37). This meant that instructors with more experience in years were less appreciative of online assessment's practical features compared to instructors with less than ten years of experience, 11-20 years of experience, and instructors with more than 30 years of experience.

According to *Table 14*, another aspect of online assessment that points out a statistically significant difference with teaching experience is the impact of online assessment on teaching and learning *(p*=.02). It was seen from the analysis that instructors with less than ten years of experience ($\bar{x}$= 3.53) had a more positive perception of the impacts of online assessment on teaching and learning compared to those with more than 30 years of experience ($\bar{x}$=3.07). This could mean that instructors with less than ten years of experience found online assessment more useful compared to those with more than 30 years of experience regarding its impacts on teaching and learning.

Table 14. Tukey's Post Hoc Analysis for Perceptions of Instructors on Online Assessment and Teaching Experience

	(I) Teaching Experience Interval	**(J) Teaching Experience Interval**	**Mean Difference**	**Std. Error**	**Sig. (p-value)**
Affective Factors	<10	11-20	.10	.08	.03
		21-30	.16	.09	
		>30*	.36	.12	
	11-20	<10	-.1	.08	
		21-30	.05	.08	
		>30*	.26	.11	
	21-30	<10	-.16	.09	
		11-20	-.05	.08	
		>30*	.20	.12	
	>30	<10*	-.36	.12	
		11-20*	-.26	11	
		21-30*	-.20	.12	
Practicality	<10	11-20	.03	.10	

		21-30	.11	.12	.04
		>30*	.43	.16	
	11-20	<10	-.03	.10	
		21-30	.08	.10	
		>30*	.40	.15	
	21-30	<10	-.11	.12	
		11-20	-.08	.10	
		>30	.31	.16	
	>30	<10*	-.43	.16	
		11-20*	-.40	15	
		21-30	-.31	.16	
Impacts on Teaching and Learning	<10	11-20	.10	.10	
		21-30	.21	.11	.02
		>30*	.45	.15	
	11-20	<10	-.10	.10	
		21-30	-.11	.10	
		>30*	.34	.14	
	21-30	<10	-.21	.11	
		11-20	-.11	.10	
		>30	.23	.15	
	>30	<10*	-.45	.15	
		11-20*	-.34	14	
		21-30	-.23	.15	

As the values in *Appendix E* illustrate, teaching experience is a significant variable in terms of scale items in affective factors, practicality aspect, and impact of online assessment on teaching and learning. Items 1 (p=.00), 3 (p=.00), and 4 (p=.00) of affective factors indicated significant differences between instructors with different experience levels. It was apparent from the mean differences that the use of computers created more stress for the instructors with more than 30 years of experience ($\bar{x}$= 3.79) compared to teachers with less than ten years of experience ($\bar{x}$= 2.97), instructors with 11-

20 years of experience ($\bar{x}$= 3.25), and those with 21-30 years of experience ($\bar{x}$=3.16). It was also clear that highly experienced teachers ($\bar{x}$= 4.44) would feel more comfortable with paper-based exams compared to those with less than ten years of experience ($\bar{x}$=3.65), those with 11-20 years of experience ($\bar{x}$=3.79) and the ones with 21-30 years of experience ($\bar{x}$=3.72). Moreover, those with more than 30 years of experience ($\bar{x}$= 3.89) felt it was more difficult to invigilate when the exam was online compared to all groups of different experience years (<10:$\bar{x}$= 2.94; 11-20:$\bar{x}$= 3.14; 21-30:$\bar{x}$= 3.36). In light of these, it can be stated that teachers with more experience years had stronger affective barriers toward online assessment compared to those with fewer experience years in general. Furthermore, *Appendix E* reveals that item 11 of practicality aspect indicates a significant difference (p=.00) between instructors with less than ten years ($\bar{x}$=3.91), instructors with 11-20 years ($\bar{x}$= 3.84), and instructors with more than 30 years of experience ($\bar{x}$= 3.27). According to the results, it can be seen that instructors with less experience believed that saving paper was a practical aspect of online assessment compared to those with more years of experience. Overall, it can be stated that the less experienced the instructors were, the more sensitive they were about the use of paper in assessment. The last item indicating significantly different results was item 28 (p=.03) in *Appendix E.* According to the table, the instructors with less than ten years of experience ($\bar{x}$=3.30) responded more positively, stating that online assessment could add value to learning more than instructors with 30 years of experience and above ($\bar{x}$=2.75) did.

4.2.4. Degree of Education

As *Table 15* shows, the degree of education is not a significant differentiating factor (p=.56) in instructors' overall perception of online assessment. The table also reveals that the relationship between the components of online assessment and the degree of completed education level indicated no significant difference. The *p*-value was .50 for affective factors, .79 for validity, .75 for practicality, .41 for reliability, .85 for security, and .10 for impacts on teaching and learning.

Table 15. Relationship Between Online Assessment Perceptions and Degree of Education

(One-way ANOVA)

Online	Degree of Education	N	Mean	Std. Deviation	F	Sig. (p-value)
Assessment Perception	BA	92	2.78	.51	.56	.56
	MA	177	2.82	.60		
	Ph.D	33	2.90	.68		
Affective Factors	BA	92	2.62	.81	.68	.50
	MA	177	2.72	.82		
	Ph.D	33	2.78	.87	.68	.50
Validity	BA	92	2.83	.68	.23	.79
	MA	177	2.78	.69		
	Ph.D	33	2.84	.72	.23	.79
Practicality	BA	92	2.68	.61	.28	.75
	MA	177	2.74	.78		
	Ph.D	33	2.69	.80	.28	.75
Security	BA	92	2.93	.65	.89	.41
	MA	177	2.95	.72		
	Ph.D	33	3.12	.86	.89	.41
Impact on Teaching and Learning	BA	92	3.31	.62	2.27	.10
	MA	177	3.38	.72		
	Ph.D	33	3.61	.77	2.27	.10

Table 16 indicates that the analysis of the scale items in each component, except for the impacts of online assessment on teaching and learning, reveals insignificant values as well. The *p*-values for the items in affective factors were between *p*=.06-.90; and *p*=.52-.93 for validity; *p*=.60-.88 for practicality, *p*=.09-.95 for reliability, *p*=.47-.94 for security. However, the p-value for item 28 of the impacts of online assessment on teaching and learning in *Table 17* indicates a significant difference, with mean score differences between teachers with a BA degree ($\bar{x}$=2.96) and teachers holding Ph.Ds ($\bar{x}$=3.51). Looking at the item, it was evident that instructors holding a Ph.D believed that online assessment added value to students' learning more than instructors holding a BA degree did.

Table 16. Relationship Between Instructor Perceptions of the Aspects of Online Assessment and Degree of Education (One-way ANOVA)

	Aspects	**Mean**	**Sig. (p-value)**	**Std. Deviation**	**F**	**Item sig. (p-values)**
Online Assessment Perception	Affective Factors	2.70	.50	.82	.82	.06-.90
	Validity	2.80	.79	.69	.69	.52-.93
	Practicality	2.72	.75	.74	.74	.60-.88
	Reliability	2.96	.41	.72	.72	.09-.88
	Security	2.32	.85	.67	.67	.47-.94
	Impacts on Teahcing and Learning	3.38	.10	.70	.70	.01-.52

Table 17. Tukey's Post Hoc Analysis for Items Indicating Statistically Significant Difference in Degree of Education

	Items	**(I) Degree**	**(J)Degree**	**Mean Difference**	**Std. Error**	**Sig. (p-value)**
Impact on	28. Online assessment can	BA	MA	-.16	.18	.01
			Ph.D*	-.54	.17	

Teaching and Learning	add value to students' language learning.	MA	BA	16	.11
			Ph.D	-.37	.18
		Ph.D	BA*	54	.18
			MA	37	.17

4.2.5. Department of Graduation

As illustrated in *Table 18*, there is no statistically significant difference in instructors' perspectives of online assessment according to their departments (*H(2)*=10.38, *p*=.11). This indicated that instructors who graduated from different departments had similar viewpoints of online assessment. In addition, according to *Table 18*, where *p*-values and *chi-square* values are indicated, it was seen that there was no significant difference between instructors' perceptions of the components of online assessment, with the exception of reliability. *Table 18* also indicates that instructors who graduated from American Culture and Literature departments ($\bar{x}$=3.28) found online assessment more reliable compared to instructors who graduated from English Language and Literature ($\bar{x}$=2.77) departments and English Language Teaching departments ($\bar{x}$=2.98).

Table 18. Relationship Between Instructor Perceptions of the Aspects of Online Assessment and Department of Graduation (Kruskal Wallis-H)

Online Assessment Perception	**Department**	**N**	**Mean Rank**	**H (chi-square)**	**Sig. (p-value)**
	English Language Teaching	188	154.16		
	English Language and Literature	74	132.59	10.88	.11
	American Culture and Literature	19	186.61		
	Other	21	162.55		

Affective Factors	English Language Teaching	188	151.69	6.15	.10
	English Language and Literature	74	136.87		
	American Culture and Literature	19	187.26		
	Other	21	169.07		
Validity	English Language Teaching	188	157.35	5.80	.12
	English Language and Literature	74	131.66		
	American Culture and Literature	19	172.42		
	Other	21	150.12		
Practicality	English Language Teaching	188	147.81	10.88	.11
	English Language and Literature	74	145.03		
	American Culture and Literature	19	201.34		
	Other	21	162.29		
Reliability	English Language Teaching	188	154.52	10.38	.01
	English Language and Literature	74	127.85*		
	American Culture and Literature	19	185.63*		
	Other	21	177.48*		
Security	English Language Teaching	188	50.95		

	English Language and Literature	74	47.25	1.05	.78
	American Culture and Literature	19	54.13		
	Other	21	69.00		
Impact on Teaching and Learning	English Language Teaching	188	157.03		
	English Language and Literature	74	139.03	3.24	.35
	American Culture and Literature	19	162.39		
	Other	21	136.10		

Table 19 indicates that an item-based analysis of the scale reveals significant differences in one item of affective factors (p=.03) and one item in reliability (p=.00). Apparently, the results of Item 1 indicated that instructors who graduated from English Language Teaching ($\bar{x}$=3.23) and English Language and Literature ($\bar{x}$=3.40) departments thought that using a computer added to the stress of exams for them more than instructors who graduated from American Culture and Literature ($\bar{x}$=2.84) and other ($\bar{x}$=2.71) departments. Moreover, instructors who graduated from English Language and Literature ($\bar{x}$=3.74) and English Language Teaching ($\bar{x}$=3.42) departments believed that paper-based assessment was more reliable than online assessments compared to instructors having graduated from American Culture and Literature ($\bar{x}$=2.94) and other ($\bar{x}$=3.14) departments.

Table 19. Tamhane's T2 Post Hoc Analysis for Items Indicating Statistically Significant Difference for the Department of Graduation

	Items	(I) Graduation Department	(J) Graduation Department	Mean Difference	Std. Error	Sig. (p-value)
Affective Factors	1. Using a computer adds to the stress of exams for teachers.	English Language Teaching	English Language and Literature	-.16	.14	.03
			Amercian Culture and Literature	.39	.25	
			Other	.52	.28	
		English Language and Literature	English Language Teaching	.16	.14	
			Amercian Culture and Literature	.56	.27	
			Other	.69	.29	
		American Culture and Literature	English Language Teaching	-.39	.25	
			English Language and Literature	-.56	.27	
			Other	.12	.36	
		Other	English Language Teaching	-.52	.28	

			English Language and Literature	-.69	.29	
			American Culture and Literature	-.12	.25	
Reliability	19. Paper based exams are fairer than online exams.	English Language Teaching	English Language and Literature	-.31	.15	.01
			Amercian Culture and Literature	.47	.27	
			Other	.28	.26	
		English Language and Literature	English Language Teaching	.31	.15	
			Amercian Culture and Literature	.79	.28	
			Other	.60	.28	
		American Culture and Literature	English Language Teaching	-.47	.27	
			English Language and Literature	-.79	.28	
			Other	-.19	.36	
		Other	English Language Teaching	-.28	.26	

English Language and Literature	-.60	.28
American Culture and Literature	.19	.36

4.2.6. Institution

As can be seen in *Table 20,* instructors' workplace is not a significant variable in their overall perceptions of online assessment. It is apparent from their mean scores that instructors working at state universities ($\bar{x}$=2.82) and private universities ($\bar{x}$=2.81) had similar perspectives of online assessment overall. Their viewpoints of the components of online assessment did not present any statistically significant difference as well, according to *Table 20*. This indicates that instructors' perspectives of different aspects of online assessment did not differ according to the institutions they worked at.

Table 20. Relationship Between Online Assessment Perceptions of Instructors and Institution (Independent Samples T-Test)

	Institution	**N**	**Mean**	**Std. Deviation**	**F**	**Sig. (p-value)**
Online Assessment Perception	State	99	2.82	.58	.35	.93
	Private	203	2.81	.58		
Affective Factors	State	99	2.72	.76	1.38	.79
	Private	203	2.69	.85		
Validity	State	99	2.81	.70	.00	.96
	Private	203	2.80	.69		
Practicality	State	99	2.73	.70	.40	.83
	Private	203	2.71	.75		
Reliability	State	99	2.88	.75		

	Private	203	3.00	.70	.44	.19
Security	State	99	2.31	.73	1.97	.84
	Private	203	2.32	.65		
Impact on Teaching and Learning	State	99	3.46	.66	1.45	.18
	Private	203	3.35	.72		

Table 21 indicates that instructors' workplace is not a significant factor for their online assessment perceptions in terms of scale items as well, except for item 8 in the validity component. According to *Table 22*, item 8 of the validity aspect indicated a significant difference, which supported the idea that instructors working at state universities ($\bar{x}$=3.60) believed that online exams also tested the IT skills of students besides the test content more than instructors working at private institutions did ($\bar{x}$=3.31).

Table 21. Relationship Between Instructor Perceptions on the Aspects of Online Assessment and Institution (Independent Samples T-Test)

Aspects of Online Assessment	**Aspects**	**Sig. (p-value)**	**Std. Error**	**F**	**Item sig. (p-values)**
	Affective Factors	.79	.10	1.38	.20-.94
	Validity	.96	.08	.00	.01-.87
	Practicality	.83	.90	.40	.44-.70
	Reliability	.19	.08	.44	.12-.87
	Security	.84	.08	1.97	.32-.98
	Impacts on Teaching and Learning	.18	.08	1.45	.06-.09

Table 22. Independent Samples T-Test for Items Indicating Statistically Significant Difference in Institution

	Items	Institution	N	Mean	Std. Deviation	F	Sig. (p-value)
Validity	8. Online exams don't just test knowledge of the subject, but IT skills as well.	State	99	3.60	.97	3.64	.01
		Private	203	3.31	1.04		

4.2.7. Administrative Duty

As can be seen in *Table 23*, instructors' perspectives of online assessment do not statistically differ according to having an administrative duty ($\bar{x}$=2.90) or not ($\bar{x}$=2.80). This could mean that regardless of being a manager, director, vice director, or coordinator, all instructors had similar perceptions of online assessment. Further analyses were carried out to see if instructors' perspectives on the different components of online assessment differed according to holding an administrative position in their institutions. According to *Table 23*, it was seen that there was no statistically significant difference in their perspectives of online assessment components according to having an administrative duty or not.

Table 23. Relationship Between Online Assessment Perceptions of Instructors and Administrative Duty (Independent Samples T-Test)

	Administrative Duty	N	Mean	Std. Deviation	F	Sig. (p-value)
Online Assessment Perception	Yes	46	2.90	.56	.15	.28
	No	256	2.80	.59		

Affective Factors	Yes	46	2.82	.82	.01	.27
	No	256	2.68	.82		
Validity	Yes	46	2.90	.65	.09	.33
	No	256	2.79	.70		
Practicality	Yes	46	2.78	.72	.27	.52
	No	256	2.71	.74		
Reliability	Yes	46	3.03	.70	.07	.48
	No	256	2.95	.72		
Security	Yes	46	2.35	.73	.56	.75
	No	256	2.31	.67		
Impact on Teaching and Learning	Yes	46	3.52	.63	2.20	.16
	No	256	3.36	.71		

Table 24 shows that the scale items in each different aspect of the online assessment do not show a statistically significant difference, except for one item in the validity component. For the other items in the scale, having an administrative duty or not was not a significant differentiating factor. According to *Table 25*, it can be stated that instructors with an administrative duty ($\bar{x}$=3.39) were more concerned about the online multiple-choice questions due to the guessing factor than the ones without an administrative duty did ($\bar{x}$=2.97).

Table 24. Relationship Between Instructor Perceptions on the Aspects of Online Assessment and Administrative Duty (Independent Samples T-Test)

Aspects of Online Assessment	**Aspects**	**Sig. (p-value)**	**Std. Error**	**F**	**Item sig. (p-values)**
	Affective Factors	.27	.13	.01	.11-.89
	Validity	.33	.11	.09	.02-.99
	Practicality	.07	.11	.27	.11-.90
	Reliability	.48	.11	.07	.08-.99

	Security	.75	.10	.56	.42-.87
	Impacts on Teaching and Learning	.16	.11	2.20	.06-.89

Table 25. Independent Samples T-Test for Items Indicating Statistically Significant Difference in Administrative Duty

	Items	**Admin. Duty**	**N**	**Mean**	**Std. Deviation**	**F**	**Sig. (p-value)**
Validity	10. Because they can guess the answer, online multiple-choice questions do not really reflect students' level of knowledge.	Yes	44	3.39	1.12	.30	.02
		No	256	2.97	1.13		

4.2.7. Office Duty

Table 26 indicates that instructors' overall perceptions of the online assessment show no significant difference with regard to having an office duty such as testing and assessment, curriculum development, or teacher training duties (p=.99). The mean scores of instructors who worked in the offices ($\bar{x}$=2.81) were similar to those who did not have any office duties ($\bar{x}$=2.81). *Table 26* also shows that a further T-test result indicated that there was no statistically significant difference in any of the components of online assessment with relation to instructors' having an office duty or not.

Table 26. Relationship Between Online Assessment Perceptions of Instructors and Office Duty (Independent Samples T-Test)

	Office Duty	N	Mean	Std. Deviation	F	Sig. (p-value)
Online Assessment Perception	Yes	96	2.81	.52	3.06	.99
	No	206	2.81	.62		
Affective Factors	Yes	96	2.69	.75	1.30	.93
	No	206	2.70	.85		
Validity	Yes	96	2.79	.64	.88	.76
	No	206	2.81	.71		
Practicality	Yes	96	2.79	.71	.44	.24
	No	206	2.68	.75		
Reliability	Yes	96	2.92	.67	7.71	.46
	No	206	2.98	.74		
Security	Yes	96	2.26	.70	.70	.30
	No	206	2.35	.66		
Impact on Teaching and Learning	Yes	96	3.43	.60	5.27	.39
	No	206	3.36	.74		

As can be seen in *Table 27,* further item analysis of the scale items indicates that having an office duty is an insignificant variable in instructors' perceptions of online assessment. However, as *Table 28* indicates, working at an office only played a significant differentiating role on item 11 of the practicality component (p=.03). Mean differences revealed that instructors who worked in offices ($\bar{x}$=3.96) were more sensitive about saving paper in online assessment compared to those without any office duties ($\bar{x}$=3.63).

Table 27. Relationship Between Instructor Perceptions of the Aspects of Online Assessment and Office Duty (Independent Samples T-Test)

Online Assessment Perception	Aspects	Sig. (p-value)	Std. Error	F	Item sig. (p-values)
	Affective Factors	.93	.10	1.30	.14-.90
	Validity	.76	.08	.88	.30-.96
	Practicality	.24	.09	.44	.03-.71
	Reliability	.46	.08	1.71	.06-.69
	Security	.30	.08	.70	.08-.91
	Impacts on Teaching and Learning	.39	.08	5.27	.08-.98

Table 28. Independent Samples T-Test for Items Indicating Statistically Significant Difference in Office Duty

	Items	Office Duty	N	Mean	Std. Deviation	F	Sig. (p-value)
Practicality	11. Online assessments use less paper, which is important to me.	Yes	96	3.92	.12	7.40	.03
		No	206	3.66	.11		

4.2.8. Self-Reported Computer Expertise

As *Table 29* indicates, instructors' self-reported computer expertise is a significant differentiating factor for instructors' online assessment perception (p=.02). According to the table, the mean differences between instructors point out that instructors who reported themselves having excellent computer skills ($\bar{x}$=2.97) had more positive perceptions of online assessment compared to those who reported themselves as having adequate

computer skills ($\bar{x}$= 2.70). *Table 29* also shows that some of the components of the online assessment indicated significant differences, while some did not. The table illustrates that instructors' perceptions of online assessment did not significantly differ in validity (p=.23), reliability (p=.22), and security (p=.57) aspects with regard to their computer expertise. Nevertheless, there were statistically significant differences in instructors' responses to affective factors (p=.00), practicality (p=.05), and in the impact of online assessment on teaching and learning (p=.00) components in relation to instructors' self-perceived computer literacy.

Table 29. Relationship Between Online Assessment Perceptions and Self-Reported Computer Expertise (One-way ANOVA)

Online	Computer Expertise	N	Mean	Std. Deviation	F	Sig. (p-value)
Assessment Perception	Adequate	65	2.70	.44*	3.97	.02
	Good	168	2.79	.59		
	Excellent	69	2.97	.65*		
Affective Factors	Adequate	65	2.51	.60*	6.81	.00
	Good	168	2.65	.82		
	Excellent	69	3.00	.94*		
Validity	Adequate	65	2.69	.57	1.45	.23
	Good	168	2.81	.68		
	Excellent	69	2.89	.80		
Practicality	Adequate	65	2.57	.58*	2.86	.05
	Good	168	2.71	.76		
	Excellent	69	2.88	.79*		
Reliability	Adequate	65	2.87	.64	1.48	.22
	Good	168	2.95	.70		
	Excellent	69	3.08	.87		
Security	Adequate	65	2.26	.61	.55	.57
	Good	168	2.32	.69		
	Excellent	69	2.38	.70		

Impact on Teaching and Learning	Adequate	65	3.30	.44*	4.96	.00
	Good	168	3.32	.59		
	Excellent	69	3.62	.65*		

Table 30 indicates that there are differences in the perceptions of instructors of the components of online assessment among different groups of instructors according to their self-perceived computer expertise. The table demonstrates that the main difference in the affective factors is between instructors who reported having excellent computer skills and those with adequate computer skills. It can be understood from *Table 29* that instructors with excellent command of computers ($\bar{x}$=3.00) perceived online assessment more positively than the instructors who thought they had good ($\bar{x}$=2.65) or adequate ($\bar{x}$=2.51) computer skills. Similarly, *Table 33* also points out the difference between instructors with adequate and excellent computer skills in the practicality aspect. Instructors who had excellent computer skills ($\bar{x}$=2.88) found online assessment more practical than the ones with adequate computer skills ($\bar{x}$=2.57). As *Table 29* indicates, self-reported computer expertise was also an effective factor in instructors' perceptions of the impacts of online assessment on teaching and learning with a *p-value* of .00. *Table 30* pinpoints the biggest differences between instructors with excellent computer skills and both of the other groups. Results revealed that instructors with an excellent command of computers ($\bar{x}$=3.62) appreciated the positive impacts of online assessment more than the instructors who reported having good ($\bar{x}$=3.32) and adequate computer skills ($\bar{x}$=3.30).

Table 30. Tukey's Post Hoc Analysis for Instructor Perceptions on Online Assessment and Self-Reported Computer Expertise Intervals

	(I) Computer Expertise	**(J) Computer Expertise**	**Mean Difference**	**Std. Error**	**Sig. (p-value)**
	Adequate	Good	-.09	.08	.02
		Excellent	-.27*	.10	

Online Assessment Perception	Good	Adequate	.09	.08	
		Excellent	-.18	.08	
	Excellent	Adequate	.27*	.10	
		Good	18	.08	
Affective Factors	Adequate	Good	-.13	.11	.00
		Excellent	-.48*	.13	
	Good	Adequate	.13	.11	
		Excellent	-.34*	.11	
	Excellent	Adequate	.48*	.13	
		Good	34*	.11	
Practicality	Adequate	Good	-.13	.10	.05
		Excellent	-.30*	.12	
	Good	Adequate	.13	.10	
		Excellent	-.16	.10	
	Excellent	Adequate	.30*	.12	
		Good	16	.10	
Impacts on Teaching and Learning	Adequate	Good	-.01	.10	.00
		Excellent	-.31*	.12	
	Good	Adequate	.01	.10	
		Excellent	-.29*	.09	
	Excellent	Adequate	.31*	.12	
		Good	29*	.09	

As *Table 31* shows, further item-based analysis of the scale indicates that there are statistically significant items in affective factors, practicality, and the impacts of online assessment on teaching and learning. The validity, reliability, and security aspects indicated no statistically significant items. Thus, self-reported computer expertise was a significant factor for the items in affective factors, practicality, and impacts on teaching and learning.

Table 31. Relationship Between Instructor Perceptions of the Aspects of Online Assessment and Self-Reported Computer Expertise (One-way ANOVA)

Online Assessment Perception	Aspects	Mean	Sig. (p-value)	Std. Deviation	F	Item sig. (p-values)
	Affective Factors	2.70	.00	.82	6.81	.00-.13
	Validity	2.80	.23	.69	1.45	.12-.92
	Practicality	2.72	.05	.74	2.86	.04-.71
	Reliability	2.96	.22	.72	1.48	.12-.66
	Security	2.32	.57	.67	.55	.06-.67
	Impacts on Teaching and Learning	3.38	.00	.70	4.96	.00-.14

Table 32 indicates that there are differences between different groups in the items of the scale as well. Instructors' points of view differed in items 3,4, and 5 in affective factors. Clearly, results for item 3 demonstrate that instructors who reported having adequate computer skills ($\bar{x}$=4.00) felt more comfortable with traditional assessment on paper compared to those with excellent computer skills ($\bar{x}$=3.55). Similarly, those with adequate ($\bar{x}$=3.52) or good ($\bar{x}$=3.26) self-reported computer skills found it hard to do relevant exam-related tasks while doing an online exam compared to those with excellent computer skills ($\bar{x}$=2.85) as item 4 suggests. Moreover, those who reported having excellent computer skills ($\bar{x}$=2.75) had fewer issues with doing exams online compared to those with good ($\bar{x}$=2.25) or adequate ($\bar{x}$=2.23) self-reported computer skills. The results demonstrate that as instructors' self-reported computer skills increased, their anxiety toward online assessment decreased. The table further presents an analysis of the individual items evaluating practicality. It is apparent in the table that instructors' perceptions statistically differed in items 11 (p=.04) and 15 (p=.03) with regard to their level of computer expertise. It was seen that instructors who had an excellent command of computers ($\bar{x}$=3.98) were more sensitive about paper use when compared to those with an adequate command of computers ($\bar{x}$=3.55) according to item 11. Also, item 15 presents

a similar result since instructors with excellent command of computers ($\bar{x}$=3.24) enjoyed the time and spatial flexibility that online assessment provides to instructors more than those with adequate self-reported computer skills ($\bar{x}$=2.73). The final point that the table indicates is in the impacts of online assessment on teaching and learning component. Item 26 indicated a statistically significant difference (p=.00), which suggested that instructors with an excellent command of computers ($\bar{x}$=3.97) agreed that the immediate feedback opportunity of online assessment added value to language learning more than instructors who reported having good ($\bar{x}$=3.55) and adequate ($\bar{x}$=3.58) command of computers.

Table 32. Tukey's Post Hoc Analysis for Items Indicating Statistically Significant Difference in Self-Reported Computer Expertise

	Items	(I) Computer Expertise	(J) Computer Expertise	Mean Difference	Std. Error	Sig. (p-value)
Affective Factors	3. I'd feel more comfortable if the exam was on paper, not online.	Adequate	Good	-.18	.15	.03
			Excellent	-.46*	.18	
		Good	Adequate	18	.15	
			Excellent	-.28	.15	
		Excellent	Adequate	.46*	.18	
			Good	.28	.15	
	4. I find it hard to invigilate/ do relevant tasks when doing and online exam.	Adequate	Good	-.26	.16	.00
			Excellent	-.66*	.19	
		Good	Adequate	.26	.16	
			Excellent	-.40*	.16	
		Excellent	Adequate	.66*	.16	
			Good	.40*	.19	
	5. I'd rather do exams on a computer than on paper because I am	Adequate	Good	-.02	.14	.00
			Excellent	-.52*	.17	
		Good	Adequate	.02	.14	
			Excellent	-.49*	.14	
		Excellent	Adequate	.52*	.17	

	used to working online.		Good	.49*	.14	
Practicality	11. Online assessments use less paper, which is important to me.	Adequate	Good	-.16	.14	.04
			Excellent	-.43*	.17	
		Good	Adequate	16	.14	
			Excellent	-.26	.14	
		Excellent	Adequate	.43*	.17	
			Good	.26	.14	
	15. Online exams are more practical than paper-based exams because they are free from time and space.	Adequate	Good	-.17	.16	.03
			Excellent	-.50*	.19	
		Good	Adequate	.17	.16	
			Excellent	-.32	.16	
		Excellent	Adequate	.50*	.19	
			Good	.32	.16	
Impacts on Teaching and Learning	26. The potential for immediate feedback with online assessment could help students learn.	Adequate	Good	.02	.12	.00
			Excellent	-.38*	.14	
		Good	Adequate	-.02	.12	
			Excellent	-.41*	.11	
		Excellent	Adequate	.38*	.14	
			Good	.41*	.11	

4.2.9. Summary

This study mainly found that teachers with a neutral perspective of online assessment greatly outnumbered teachers with positive and negative perspectives. The study also revealed that whereas the greatest concern of instructors was the security aspect of online assessment, they seemed to appreciate the positive impact of online assessment on teaching and learning to some degree. The second major finding of this study was that factors such as gender, age, years of teaching experience, and computer expertise had an impact on how instructors perceived online assessment. On the other hand, degree of education completed, department of graduation, workplace, and having an administrative or office duty were not factors affecting instructors' perspectives significantly, although there were statistically significant differences in the scale items for almost all variables. A detailed summary of the findings can be found in *Table 33* below.

Table 33. Summary of the Results

Research Question	**Results**
RQ1- How do EFL instructors perceive using online testing and assessment?	Mean average of instructors' perspectives toward online assessment was 2.81 over 5.00. This indicates that instructors mostly had a neutral perspective of online assessment.
RQ2. Is there a significant difference between the perceptions of EFL instructors toward online assessment and certain variables?	
a. Gender	Significant variable (p=.00) Males had a more positive perspective of online assessment.
b. Age	Significant variable (p=.05) Instructors who were younger had a more positive perspective of online assessment.

c. Teaching experience in years	Significant variable (p=.03) Instructors with fewer years of experience had a more positive perspective of online assessment.
d. Degree of education completed	No overall significance was found. 1 item significantly differed in the category of impacts on teaching and learning.
e. Department of graduation	No overall significance was found. 1 item significantly differed in the category of reliability. 1 item significantly differed in the category of affective factors.
f. Institution	No overall significance was found. 1 item significantly differed in the category of reliability.
g. Administrative Duty	No overall significance was found. 1 item significantly differed in the category of validity.
h. Office Duty	No overall significance was found. 1 item significantly differed in the category of practicality.
i. Self-reported computer expertise	Significant variable (p=.02) Instructors with excellent self-reported computer expertise had a more positive perspective of online assessment.

5. CONCLUSION AND DISCUSSION

The conclusion and discussion part consists of four sections. The first section presents the findings of this study. The second section discusses the findings with regard to relevant literature. The third section gives recommendations for the practical use of these findings, and the last section gives recommendations for further research.

5.1. Conclusions

This study aims at investigating teachers' perceptions of online testing and assessment. Given that the Internet and computers have inevitably been integrated into many aspects of education because of advancing technology (Momeni, 2022), it is of utmost importance to discover how teachers perceive online assessment. To this end, instructors' overall perceptions of online assessment are identified, and, afterward, their overall perceptions of different components of online assessment are studied; finally, the relationships between their perceptions and certain variables are analyzed. This study concludes that most instructors hold a neutral perception of online assessment. It is apparent from instructors' responses that although a small minority seems to appreciate the advantages of online assessment in terms of practicality and pedagogy, the vast majority seems to have serious concerns regarding anxiety, difficulty, reliability, and security. Secondly, it is found that instructors' overall perspectives of online assessment differ significantly according to gender, age, teaching experience, and their self-perceived computer expertise. However, degree of completed education, graduation department, workplace, and having an office or administrative duty seem to be insignificant factors affecting instructors' perspectives of online assessment.

According to the results of the study, gender is a statistically significant variable affecting instructors' perspectives of online assessment. It is found that male instructors have a more positive overall perception of online assessment compared to females in this context. They have fewer anxiety issues and feel more comfortable compared to females during an online assessment. Another critical finding of this study is age being an important differentiating variable in instructors' perceptions of online assessment. Instructors younger than 40 years old have more positive perspectives of online assessment compared to those 50 years old and older. Naturally, the results and findings of age and teaching experience are compatible with each other since these two variables

are demographically in close connection with each other. As said earlier, teaching experience is effective in creating statistically significant differences between different experience groups. In general, it is seen that instructors with less experience have more positive perspectives with regard to online assessment compared to teachers with long years of experience, reflecting the results of the age variable. The final variable that affects instructors' perceptions of online assessment is self-reported computer expertise. Instructors with higher self-reported computer expertise have a more positive perception of online assessment than those who report themselves as adequate or good users of computers. On the other hand, the highest degree of completed education, department of graduation, the institution of work, having an administrative duty or not, and having an office duty are not significant variables in instructors' perceptions of online assessment. However, significant differences are observed with regard to scale items for these variables, as well.

5.2. Implications and Discussion

Researching and analyzing teacher perceptions of assessment is crucial for two main reasons. First of all, as one of the most important components of teaching and learning practices, assessment is necessary to document whether learning objectives have been met. Without assessment, it might not be possible to make educated decisions about many issues including curriculum, teaching practices, learner behavior, and learning goals. Since assessment is crucial to that extent, it is important to understand the factors that impact or control the driving forces behind it, one of which is teacher perceptions and beliefs (Teo & van Schaik, 2012). Second, research suggests that teachers' beliefs and perspectives significantly affect assessment practices, which shapes teaching and learning. For this, sufficient time and effort should be devoted to investigating these perceptions, beliefs, and perspectives and declaring them openly not only with publicly accepted general statements but also with the results of more personal investigations such as research on teacher perceptions (Brown, 2002; Teo & van Schaik, 2012). Thus, teacher perception of assessment is paramount to investigate with proper research.

This study shares similar results with many studies in existing literature in terms of many aspects. Many existing studies in the literature indicate that teachers have profound concerns and worries regarding security issues. Rogers (2006), Mellar et al.

(2018), Meccawy et al. (2021), Alghammas (2020), and Sa'di (2021) are some of the researchers that conclude that teachers have serious security and academic integrity concerns for online assessment. This study indicates that the same concerns are shared by teachers as seen in instructors' responses to security component, especially to item 23 stating that "it is easier to cheat on online exams than with paper-based exams", both of which indicate that instructors' greatest concern in online assessment is security. Moreover, this study also shares common findings with Rollim and Isaias (2018), as both studies indicate that teachers lack trust in the system of online assessments as seen in the mean score of item 24 which states that "the online exam system is vulnerable to hackers". Another similarity of results between this study and other studies such as Mirza (2021), Zhang et al. (2021), and Astiandani and Anam (2020) is that instructors feel anxious during online assessment due to many reasons. This study also concludes that instructors' stress levels increases during online assessment, according to their responses to the first five items. Another similar finding is on technical barriers. The current study shares similar results with Nowroozi (2021) and Joshi (2020) as all conclude that instructors face technical barriers during online assessment. In this study, instructors' perspectives on technical issues are apparent in their responses to item 12 stating that "technical problems make online exams impractical". Furthermore, similar to Yulianto and Mujtahid (2021) who find in their research that teachers feel online assessment as less effective compared to paper-based exams, this study reveals that the majority of teachers would feel more comfortable with paper-based exams, and they would prefer online exams less than paper-based exams as item 3 and 5 indicate. Moreover, they find paper-based exams fairer than online exams according to item 19, which indicates that the instructors find online assessment less efficient compared to traditional assessment. Another prominent research that shares common findings with this study belongs to Küppers and Schroeder (2020), stating that teachers who are younger and have better computer literacy have more positive perspectives of online assessment. The current study is in perfect line with Küppers and Schroeder (2020)'s study since it has found exactly the same results as can be seen in the results section.

There are also contrasting results of this study with others in literature with regard to many findings. First, this study shares contradictory results with those claiming that

instructors have positive perspectives of online assessment with less anxiety when the assessment is online as in Baleni (2015)'s research. In the current study, on the other hand, instructors seem to have neutral perspectives towards online assessment, with a serious level of anxiety, as can be seen in the mean scores of instructors in affective factors. Moreover, this study has contradictory points with Abduh (2021), who concludes that gender is not a statistically significant differentiating factor in teacher perceptions toward online assessment. In the current study, gender is found to be a significant factor (p=.00), with males having a more positive perception of online assessment than females. Another research this study has contradictory results with is Öz's (2014b) study, which concludes that female instructors feel less anxious during an online assessment. Whereas in this study, it is found that female instructors are more anxious during online assessment compared to males, according to their responses in the affective factors component. Other studies in the literature that have different results with this study are Chien et al (2014)'s and Fageah (2015)'s study both of which reveal that teachers have positive attitudes toward online assessment. However, the results of this study reveal that instructors had a neutral perception toward online assessment. In sum, the current study has contradictory results with other studies in literature, some of which found that instructors have less anxiety during online assessment, they have positive attitudes toward it and gender is either not a significant variable in teachers' online assessment perceptions or even females are less anxious during online assessment.

There are also studies that are in partial agreement with the results of this study. First of all, this study partially complies with Küppers and Schroeder (2020)'s study and Fitriyah and Jannah (2021)'s study, both of which reveal that instructors have positive perspectives toward online assessment, but had security concerns in mind. This study exhibits similar findings to the aforementioned studies, as security emerges as the primary concern based on the responses of participating instructors, while having neutral perspectives toward online assessment. Another study that aligns, to some extent, to this study is Asma (2021)'s research, which reveals that teachers have positive attitudes towards online assessment owing to its benefits such as being flexible and eco-friendly, but have concerns toward it because of screen fatigue and cheating issues. The current study also unveils instructors' concerns on screen fatigue and cheating, apparent in items

13 and 23. However, it has contradicting results with Asma (2021) in terms of the overall attitude of instructors as this study concludes that instructors have a netural perception of online assessment.

5.3. Practical Recommendations

Several practical recommendations can be made in light of the findings of this study for researchers, instructors, and for decision-makers. First and foremost, it must be noted that instructors' perspectives of online assessment should be researched in as many contexts as possible since they are actually the "agents of the assessment process" (Harlen, 1996 as cited in Shim, 2009). Being the core users of the system, they contribute greatly to the decision-making process in relation to assessment, teaching, learning, policies, and curriculum. In other words, how teachers perceive online assessment seriously affects the way they implement online assessment practices in and outside the class. Therefore, as teachers' beliefs and perceptions greatly influence their practices, a considerable amount of research should be done to understand their perspectives on all types of assessment, including online assessment, and the factors that influence their perceptions (Shim, 2009). Secondly, because the use of computers and the Internet has been increasing rapidly in recent years in the language teaching context, online assessment should be the focus of research. It is necessary to understand the dynamics, principles, designs, and pedagogical implications of online assessment to be able to fully utilize the unprecedented potential of online assessment for students' learning and teachers' job (Stödberg, 2015).

Teachers can also benefit from practical recommendations in light of this research. First of all, since it was found that teachers have a neutral perspective of online assessment with many issues in mind such as security and validity concerns, less appreciation of flexibility of time and space, non-acceptance of immediate feedback chances, and disapproval of potential positive contributions to class learning, it is necessary to raise awareness on such issues with relevant input. Thus, it is essential to help instructors understand and appreciate the nature of online assessment with to-the-point workshops or training sessions. Secondly, this research puts forward two important conclusions: teachers' anxiety level increases greatly with age and less computer expertise. Accepting that online assessment might be an inevitable component of assessment, especially in the higher education context, it might be necessary to overcome these issues as much as

possible. To this end, teachers, regardless of their age, must be given necessary computer or online assessment system training, and this training should be updated at regular intervals to keep the teachers' skills up-to-date.

As for decision makers such as school administrators, testing office members, and test writers, several recommendations can be put forward. First of all, as it is evident that teachers' overall perception of online assessment is neutral, security being the least positive aspect of online assessment according to teachers' perspectives, decision-makers can improve the security of the systems against both cheating and hackers. To this end, several measures such as using a secure browser technology, exam recording, auto and live proctoring methods can be taken. Moreover, exam data should be kept in well-protected virtual areas so hackers cannot enter the system. In order to ensure security and prevent technology-related misconduct or failures, continuous IT support should be provided before, during, and after the exam. Secondly, the findings of this study suggest that instructors with an adequate level of computer expertise found online assessment less positive compared to the ones with excellent computer skills. Decision makers should be aware of this problem and provide the teachers with the necessary computer training so that the instructors feel more comfortable during online exams. Finally, the decision-makers should be aware of the benefits and opportunities that online assessment might provide and promote online assessment in their institutions by giving relevant information and training to the teachers working at their institutions. As online assessment will probably be much more common in the future thanks to its ease of use, efficient administration, ease of grading, and grade announcement; its flexible nature of time and space; its prospect to give immediate feedback to the student, and the teacher, institutions, and decision-makers need to be ready to equip their instructors with essential information and skills. Moreover, they need to be ready to make necessary technological innovations and install the required equipment to adapt to 21st-century assessment methods.

5.4. Recommendations for Further Research

The further research recommendations include recommendations in terms of research methods, research participants, the scope of the research, and the theories to be tested. As mentioned earlier, research investigating teachers' perspectives of online assessment is limited in number, and most of the time, they are limited to qualitative

studies. This research in the Turkish higher education context has been one of the firsts investigating the issue in a quantitative research method. Therefore, more quantitative studies investigating the issue would be better to reach more comprehensive and conclusive results. In addition, supporting the quantitative research with qualitative data such as interviews with teachers will help the results become more comprehensive and informative.

Moreover, data collection can be done from multiple resources. Although the data collected from instructors is meaningful, seeing the issue from the eyes of the students is crucial to understand the issue of online assessment in a better and clearer way as they are also at the core of assessment along with teachers. Therefore, conducting a study with a similar focus on students and teachers is essential. Moreover, in order to view the issue from the administrators' and testing office members' viewpoints, it could be wiser to involve them in the process. Involving as many relevant parties as possible will allow data collection in a multifaceted way, making data triangulation possible multiple times. Thus, more sound and meaningful decisions can be made accordingly. Expanding the scope of further studies would be wise to come to rational conclusions and make meaningful decisions. The scope of the present study is limited to university instructors only. Since online assessment is not used only in tertiary educational institutions, it is necessary to widen the scope of this study to high schools and, if applicable, other institutions. If possible, it would even be better to expand the scope of the study to both the students and teachers of different institutions so that educated decisions can be made more accurately.

Finally, as online assessment systems are relatively new systems of assessment and testing, it is highly important to see the level of acceptance of it by its users. Understanding which users will accept technological tools and systems is evaluated by Technology Acceptance Model (Venkatesh & Davis, 2000). Since Technology Acceptance Model (TAM) is efficiently capable of explaining if a technological improvement is or will be accepted by the users and will continue to be used or not, it is necessary to see the acceptance level of both instructors and students. This way, it could be easier to make generalizations about the public before actually using a technological advancement. Explaining online assessment with TAM would also help researchers to understand the factors and systems that affect the level of acceptance by the relevant

parties. Thus, studying online assessment in light of TAM would be very helpful. In sum, online assessment is a very important and recent technological improvement that has impacted the area of language teaching to a great extent. To this end, more research projects should be carried out to disclose areas worth investigating. With further research, it might be possible to shed light on the underresearched areas of online assessment.

REFERENCES

Abduh, M. Y. M. (2021). Full-time online assessment during COVID-19 lockdown: EFL teachers' perceptions. *Asian EFL Journal Research Articles*, *28*(1.1), 26-46.

Adanir, G. A., Ismailova, R., Omuraliev, A., & Muhametjanova, G. (2020). Learners' perceptions of online exams: A comparative study in Turkey and Kyrgyzstan. *International Review of Research in Open and Distance Learning*, *21*(3), 1–17. https://doi.org/10.19173/irrodl.v21i3.4679

Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Psychology*, *22*(1), 453–474.

Al-alak, B., & Alnawas, I. (2011). Measuring the acceptance and adoption of E-learning by academic staff. *An International Journal*, *3*(2), 201–221.

Alderson, C. J. (1981). Introduction. In C. J. Alderson & A. Huges (Eds.), *Issues in language testing* (pp. 5–9). The British Council.

Alghammas, A. (2020). Online language assessment during the COVID-19 pandemic: University faculty members' perceptions and practices. *Asian EFL Journal*, *27*(44), 169–195.

Alkış, N., & Özkan, S. (2010). Work in progress - a modified technology acceptance model for e-assessment: Intentions of engineering students to use web-based assesment tools. *40th ASEE/IEEE Frontiers in Education Conference*.

Alruwais, N., Wills, G., & Wald, M. (2018). Advantages and challenges of using e-assessment. *International Journal of Information and Education Technology*, *8*(1), 34–37. https://doi.org/10.18178/ijiet.2018.8.1.1008

Alsadoon, H. (2017). Students' perceptions of e-assessment at Saudi Electronic University. *TOJET: The Turkish Online Journal of Educational Technology*, *16*(1), 147–153.

Appiah, M., & van Tonder, F. (2018). E-assessment in higher education: A review. *International Journal of Business Management and Economic Research*, *9*(6), 1454–1460.

Arnold, I. J. M. (2016). Cheating at online formative tests: Does it pay off? *Internet and Higher Education*, *29*, 98–106. https://doi.org/10.1016/j.iheduc.2016.02.001

Asma, K. (2021). *Examining online testing practices in foreign language education through the perceptions of instructors and students: A case study* [Master's Thesis]. Akdeniz University.

Astiandani, F. R., & Anam, S. (2021). EFL teachers' perceptions towards the implementation of online formative assessment amidst the COVID- 19 pandemic. *ELT Worldwide: Journal of English Language Teaching*, *8*(2), 269–277.

Aydın, S. (2004). *The efficiency of computer on testing writing skills* [Ph. D. Dissertation]. Atatürk University.

Babo, R., Amorim, J. L., & Suhonen, J. (2018). E-assessment with multiple choice questions: A qualitative study of teachers' opinions and experience regarding the new assessment strategy. *International Journal of Learning Technology*, *13*(3), 220–248.

Babo, R., Babo, L., Suhonen, J., & Tukiainen, M. (2020). E-assessment with multiple-choice questions: A 5 year study of students' opinions and experience. *Journal of Information Technology Education: Innovations in Practice*, *19*, 1–29. https://doi.org/10.28945/4491

Bachman, L. F. (1990). *Fundamental considerations in language testing* (3rd ed.). Oxford.

Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge University Press.

Bachman, L., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.

Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.

Bahar, M. (2014). Üniversite öğrencileri için e-değerlendirme tutum ölçeğinin geliştirilmesi [Devising a scale for university students' e-assessment attitude]. *E-International Journal of Educational Research*, *5*(2), 38–53.

Bailey, S., Hendricks, S., & Applewhite Stephen, S. F. (2015). Student perspectives of assessment strategies in online courses. *Journal of Interactive Online Learning*, 13(3), 112-125.

Balaman, F., & Tiryaki, H. (2021). The opinions of teachers about compulsory distance education due to Corona Virus (Covid-19). *Journal of the Human and Social Sciences Researches*, *10*(1), 52–84.

Baleni, Z. G. (2015). Online formative assessment in higher education: Its pros and cons. *Journall of E-Learning , 13*(4), 228–236.

Benson, R., & Brack, C. (2010). *Online learning and assessment in higher education: A planning guide* (1st ed., Vol. 1). Chandos Publishing.

Bernik, I., & Jereb, E. (2006). Students' readiness for electronic examinations. *Proceedings of the 5th WSEAS International Conference on Education and Educational Technology*, 142-145.

Biggs, J., & Tang, C. (2007). Using constructive alignment in outcomes-based teaching and learning teaching for quality learning at university (3rd ed.). Maidenhead: Open University Press.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Kappan Magazine*, *92*(1), 81–91.

Boitshwarelo, B., Reedy, A. K., & Billany, T. (2017). Envisioning the use of online tests in assessing twenty-first century learning: A literature review. *Research and Practice in Technology Enhanced Learning*, *12*(16), 1–16. https://doi.org/10.1186/s41039-017-0055-7

Brown, G. T. L. (2002). *Teachers' conceptions of assessment* [Doctoral Thesis]. The University of Auckland.

Brown, H. D. (2003). *Language assessment: Principles and classroom practices* (1st ed., Vol. 1). Pearson.

Brown, J. D. (2012). Classical test theory. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (1st ed., Vol. 1, pp. 323–335). Routledge.

Butler-Henderson, K., & Crawford, J. (2020). A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Computers & Education*, 159(104024), 1-12.

Çağlayan, E. (2021). E-assessment in language teaching. In S. İnal & O. Tunaboylu (Eds.), *Language assessment theory with practice* (1st ed., Vol. 1, pp. 263–276). Nobel.

Carr, N. (2011). *Designing and analyzing language tests*. Oxford .

Chapelle, C. A., & Douglas, D. (2007). *Assessing language through computer technology* (C. Alderson & L. Bachman, Eds.; 2nd ed.). Cambridge University Press.

Chapelle, C., & Voss, E. (2008). Utilizing technology in language assessment. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment* (3rd ed., Vol. 1, pp. 149–162). Springer.

Chien, S. P., Wu, H. K., & Hsu, Y. S. (2014). An investigation of teachers' beliefs and their use of technology-based assessments. *Computers in Human Behavior*, *31*(1), 198–210. https://doi.org/10.1016/j.chb.2013.10.037

Davis, Fred. D. J. (1985). *A technology acceptance model for empirically testing new end user information systems: Theory and results* [PhD Dissertation]. Massachusetts Institute of Technology.

Dermo, J. (2009). E-assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology*, *40*(2), 203–214. https://doi.org/10.1111/j.1467-8535.2008.00915.x

Dornyei, Z. (2003). *Questionnaires in second language research: Construction, administration, and processing* (1st ed., Vol. 1). Lawrence Erlbaum Associates, Publishers.

Douglas, D. (2010). *Understanding language testing* (B. Comrie & G. Corbett, Eds.; 1st ed.). Routledge.

Fageeh, A. I. (2015). EFL student and faculty perceptions of and attitudes towards online testing in the medium of Blackboard: Promises and challenges. *Jaltcall Journal*, *11*(1), 41–62.

Farhady, H. (2021b). Qualities of a language test: Reliability. In S. İnal & O. Tunaboylu (Eds.), *Language assessment theory with practice* (1st ed., Vol. 1, pp. 37–56). Nobel.

Ferretti, F., Santi, G. R. P., Zozzo, A. del, Garzetti, M., & Bolondi, G. (2021). Assessment practices and beliefs: Teachers' perspectives on assessment during long distance learning. *Education Sciences*, *11*(6), 1–17. https://doi.org/10.3390/educsci11060264

Fitriyah, I., & Jannah, M. (2021). Online assessment effect in EFL classroom: An investigation on students and teachers' perceptions. *IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics)*, *5*(2), 265. https://doi.org/10.21093/ijeltal.v5i2.709

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book* (C. Candlin & R. Carter, Eds.; 1st ed.). Routledge Applied Linguistics.

Fulcher, G., & Davidson, F. (2012). *The Routledge handbook of language testing* (G. Fulcher & F. Davidson, Eds.; 1st ed., Vol. 1). Routledge.

Fulcher, Glenn. (2010). *Practical language testing*. Hodder Education.

Gamage, K. A. A., de Silva, E. K., & Gunawardhana, N. (2020). Online delivery and assessment during COVID-19: Safeguarding academic integrity. *Education Sciences*, *10*(11), 1–24. https://doi.org/10.3390/educsci10110301

Gass, S., Mackey, A., & Ross-Feldman, L. (2005). Task-based interactions in classroom and laboratory settings. *Language Learning*, 55(4), 575–611. https://doi.org/10.1111/j.0023-8333.2005.00318.x

Gaytan, J., & McEwen, B. C. (2007). Effective online instructional and assessment strategies. *International Journal of Phytoremediation*, *21*(1), 117–132. https://doi.org/10.1080/08923640701341653

Ghanbari, N., & Nowroozi, S. (2021). The practice of online assessment in an EFL context amidst COVID-19 pandemic: Views from teachers. *Language Testing in Asia*, *11*(27), 1–18. https://doi.org/10.1186/s40468-021-00143-4

Göktürk Sağlam, A. L. (n.d.). Basic concepts in assessment . In E. Solak (Ed.), *Assessment in language teaching* (pp. 9–23). Pelikan .

Green, A. (2014). *Exploring language assessment and testing: language in action* (R. Carter & G. Cook, Eds.; Vol. 1). Routledge Introductions to Applied Linguistics.

Hadijah, H., Pratolo, B. W., & Rondiyah, R. (2020). Interactive game "Kahoot!" as the media of students' vocabulary assessment. *Journal on English as a Foreign Language,* 10(1), 87–105. https://doi.org/10.23971/jefl.v10i1.1670

Hichour, H. (2022). Teachers' experience in e-assessment: Case study of EFL teachers in Algerian universities. *Arab World English Journal*, *13*(1), 450–461.

Hughes, A. (1989). *Testing for language teachers* (2nd ed.). Cambridge.

Ilgaz, H., & Afacan Adanır, G. (2020). Providing online exams for online learners: Does it really matter for them? *Education and Information Technologies, 25*(2), 1255–1269. https://doi.org/10.1007/s10639-019-10020-6

İnal, S., & Kazazoğlu, S. (2021). Perspectives on assessment and feedback. In S. İnal & O. Tunaboylu (Eds.), *Language assessment theory with practice* (1st ed., Vol. 1). Nobel Publishing.

Joint Information Systems Committee. (2007). *Effective practice with e-assessment*. www.jisc.ac.uk

Joshi, A., Vinay, M., & Bhaskar, P. (2020). Impact of coronavirus pandemic on the Indian education sector: Perspectives of teachers on online teaching and assessments. *Interactive Technology and Smart Education*, *18*(2), 205–226. https://doi.org/10.1108/ITSE-06-2020-0087

Khan, S., & Khan, R. A. (2019). Online assessments: Exploring perspectives of university students. *Education and Information Technologies*, *24*(1), 661–677. https://doi.org/10.1007/s10639-018-9797-0

Korkmaz, G. (2019). *Perceptions and attitudes of EFL learners and their instructors towards online speaking assessment* [Master's Thesis]. Pamukkale Universitesi.

Kömür, Ş., & Erol, İ. K. (2021). Qualities of a language test. In O. Tunaboylu & S. İnal (Eds.), *Language assessment theory with practice* (1st ed., Vol. 1, pp. 75–98). Nobel.

Kundu, A., & Bej, T. (2021). Experiencing e-assessment during COVID-19: An analysis of Indian students' perception. *Higher Education Evaluation and Development, 15*(2), 114–134. https://doi.org/10.1108/heed-03-2021-0032

Küppers, B., & Schroeder, U. (2020). Teacher's perspective on e-assessment: A case study from Germany. *CSEDU 2020 - Proceedings of the 12th International Conference on Computer Supported Education*, *1*, 503–509. https://doi.org/10.5220/0009578105030509

Linn, R. L., & Miller, D. M. (2005). *Measurement and assessment in teaching* (9th ed.). Pearson Education.

Long, A. Y., Shin, S.-Y., Geeslin, K., & Willis, E. W. (2018). Does the test work? Evaluating a web-based language placement test. *Language Learning & Technology*, *22*(1), 137–156. https://dx.doi.org/10125/44585

Madden, T. J., Ellen, P. S., & Ajzen, I. (1992). A comparison of the theory of planned behavior and the theory of reasoned action. *Personality and Social Psychology Bulletin*, *18*(1), 3–9.

Madsen, Harold. S. (1983). *Techniques in testing*. Oxford University Press.

Mahbub, M. A. (2020). An investigation into undergraduate students' perception of Kahoot mediated e-assessment. *Journal of English Education and Linguistics Studies*, *7*(2), 269–296. https://doi.org/10.30762/jeels.v7i2.2060

Meccawy, Z., Meccawy, M., & Alsobhi, A. (2021). Assessment in 'survival mode': student and faculty perceptions of online assessment practices in HE during Covid-19 pandemic. *International Journal for Educational Integrity*, *17*(1), 1-24. https://doi.org/10.1007/s40979-021-00083-9

Mede, E., & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Dil Dergisi*, *168*(1), 43–60.

Mellar, H., Peytcheva-Forsyth, R., Kocdar, S., Karadeniz, A., & Yovkova, B. (2018). Addressing cheating in e-assessment using student authentication and authorship checking systems: Teachers' perspectives. *International Journal for Educational Integrity*, *14*(2), 1–21. https://doi.org/10.1007/s40979-018-0025-x

Mirza, H. S. (2021). University teachers' perceptions of online assessment during the Covid-19 pandemic in Lebanon. *American Academic & Scholarly Research Journal*, *13*(1), 11–24.

Momeni, A. (2022). Online assessment in times of COVID-19 lockdown: Iranian EFL teachers' perceptions. *International Journal of Language Testing*, 12(2), 1-24.

Morrow, K. (1981). Communicative language testing: revolution or evolution . In C. J. Alderson & A. Hughes (Eds.), *Issues in language testing* (1st ed., pp. 9–26). The British Council.

Muñoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *Profile*, *14*(1), 143–158.

Muñoz-Leiva, F., Sánchez-Fernández, J., Montoro-Ríos, F., & Ibáñez-Zapata, J. Á. (2010). Improving the response rate and quality in Web-based surveys through the personalization and frequency of reminder mailings. Quality and Quantity, 44(5), 1037–1052. https://doi.org/10.1007/s11135-009-9256-5

Ockey, G. J. (2012). Item response theory. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (1st ed., Vol. 1, pp. 336–347). Routledge.

Öz, H. (2014a). Turkish teachers' practices of assessment for learning in the English as a foreign language classroom. *Journal of Language Teaching and Research*, *5*(4), 775–785. https://doi.org/10.4304/jltr.5.4.775-785

Öz, H. (2014b). Pre-service English teachers' perceptions of web-based assessment in a pedagogical content knowledge course. *Procedia - Social and Behavioral Sciences*, *141*(1), 45–58. https://doi.org/10.1016/j.sbspro.2014.05.010

Pehlivan Şişman, E., & Büyükkarcı, K. (2019). A review of foreign language teachers' assessment literacy. *Sakarya University Journal of Education*, *9*(3), 628–650. https://doi.org/10.19126/suje.621319

Prensky, M. (2010). *Teaching digital natives: Partnering for real learning* (1st ed.). Corvin.

Rea-Dickins, P. (2004). Editorial understanding teachers as agents of assessment. *Language Testing*, *21*(3), 249–258.

Richards, J. C. (2001). *Curriculum development in language teaching* (Vol. 1). Cambridge Language Education.

Ridgway, J. & S. M. S., McCusker, S., & Pead, D. (2004). *Literature review of e-assessment*. Project Report, Futurelab, Bristol.

Riera Guasp, J., Ardid Ramírez, M., Gómez-Tejedor, J. A., Vidaurre Garayo, A. J., & Meseguer-Dueñas, J. M. (2018). Students' perception of auto-scored online exams in blended assessment: Feedback for improvement. *Educacion XX1*, 21(2), 79–103. https://doi.org/10.5944/educXX1.19559

Rogers, C. F. (2006). Faculty perceptions about e-cheating during online testing. *CCSC: Southeastern Conference*, 206–212.

Rolim, C., & Isaias, P. (2018). Examining the use of e-assessment in higher education: teachers and students' viewpoints. *British Journal of Educational Technology*, *0*(0), 1–16. https://doi.org/10.1111/bjet.12669

Sa'di, R. A., Abdelraziq, A., & Sharadgah, , Talha A. (2021). E-assessment at Jordan's universities in the time of the COVID-19 lockdown: Challenges and solutions. *Arab World English Journal*, *Special Issue*(1), 37–54. https://doi.org/10.24093/awej/covid.3

Sahinkarakas, S. (2012). The role of teaching experience on teachers' perceptions of language assessment. *Procedia - Social and Behavioral Sciences*, *47*, 1787–1792. https://doi.org/10.1016/j.sbspro.2012.06.901

Şanli, R. (2003). *Students' perceptions about online assessment: A case study* [Master's Thesis]. Middle East technical University.

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Sage Journals, 30*(3), 1–12.

Seliger, W. H., & Shohamy, E. (1989). *Second language research methods*. Oxford University Press.

Sevilen, Ç. (2021). *Exploring the assessment literacy of EFL instructors and students' perception about assessment in English preparatory program* [Master's Thesis]. Bahcesehir University.

Shim, K. N. (2009). *An investigation into teachers' perceptions of classroom based assessment of English as a foreign language in Korean primary education* [Ph. D. Dissertation]. University of Exeter.

Shohamy, E., Inbar- Lourie, O., & Poehner, M. E. (2008). *Investigating assessment perceptions and practices in the advanced foreign language classroom (Report No. 1108)*. https://www.researchgate.net/publication/268304619

Shraim, K. (2019). Online examination practices in higher educations institutions: Learners' perspectives. *Turkish Online Journal of Distance Education*, *20*(4), 185–196.

Spratt, M. (2005). Washback and the classroom: the implications for teaching and learning of studies of washback from exams. *Research*, *9*(1), 5–29.

Szabó, G. (2008). *Applying item response theory in language test item banking* (7th ed., Vol. 10). Peter Lang.

Teo, T., & van Schaik, P. (2012). Understanding the intention to use technology by preservice teachers: An empirical test of competing theoretical models. *International Journal of Human-Computer Interaction*, *28*(3), 178–188. https://doi.org/10.1080/10447318.2011.581892

Tim McNamara. (2000). *Language testing*. Oxford.

Trotter, E. (2006). Student perceptions of continuous summative assessment. *Assessment and Evaluation in Higher Education*, *31*(5), 505–521. https://doi.org/10.1080/02602930600679506

Tunaboylu, O. (2021). Types of tests. In S. İnal & O. Tunaboylu (Eds.), *Language assessment theory with practice* (1st ed., Vol. 1, pp. 19–35). Nobel.

Uysal, H. H. (2021). The qualities of a language test: Validity. In S. İnal & O. Tunaboylu (Eds.), *Language assessment theory with practice* (1st ed., Vol. 1, pp. 57–71). Nobel.

van Merriënboer, J. J. G., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development*, *53*(1), 5–13.

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177.

Venkatesh, V., & Davis, F. D. (2000). Theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Walker, D. J. (2007). Principles of good online assessment design. *International Online Conference on Assessment Design for Learner Responsibility*. http://www.reap.ac.uk

Wright, K. B. (2005). Researching internet-based populations: advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, 10(3).

Xerri, D., & Vella Briffa, P. (2018). Teacher involvement in high-stakes language testing. In *Teacher Involvement in High-Stakes Language Testing*. Springer International Publishing. https://doi.org/10.1007/978-3-319-77177-9

Yavuz Kırık, M. (2008). *Yabancı dil olarak İngilizce öğretmenlerinin ölçme değerlendirme bağlamında tutum ve yaklaşımları. [Approaches and attitudes of EFL teachers in the EFL testing and assessment context]* [Doctoral Thesis]. İstanbul University.

Yulianto, D., & Mujtahid, N. M. (2021). EFL teachers' perspectives and their practices. *Journal of English Teaching*, *7*(2), 229–242. https://doi.org/10.33541/jet.v7i2.2770

Zhang, C., Yan, X., & Wang, J. (2021). EFL teachers' online assessment practices during the COVID-19 pandemic: Changes and mediating Factors. *Asia-Pacific Education Researcher*, *30*(6), 499–507. https://doi.org/10.1007/s40299-021-00589-3

APPENDIX A.

Teacher Perceptions Toward Online Assessment Survey

Dear Colleague, The aim of this study is to explore and identify teacher perceptions toward online assessment in the English as a foreign language context in Turkey. You are kindly asked to spare 10 minutes to answer 30 questions in this questionnaire. I expect you to complete the survey with your honest responses. This is an anonymous questionnaire, therefore your name will not be used but the data you provided can be used for academic purposes. You can contact me at for any questions or concerns. Thank you very much for your participation.

İrem Gedil

MA Student at İstanbul Medeniyet University

Instructor at Sabancı University

Demographic Information	
Gender	Male
	Female
Age	
I have been teaching English for…years	
I graduated from … department.	English Language Teaching
	English Language and Literature
	American Culture and Literature
	Other
The Highest Degree Received	BA
	MA
	PhD

I work at a…	State University
	Private/ Foundation University
I have an administrative duty (school head, principle, manager, vice principle, coordinator etc.)	Yes
	No
I have an office duty (testing office, materials development unit, curriculum unit etc.)	Yes
	No
My computer skills are…	Poor
	Adequate
	Good
	Excellent

Survey Questions	**1**	**2**	**3**	**4**	**5**
1. Using a computer adds to the stress of exams for teachers.					
2. I expect computers to be used as part of regular assessment at university.					
3. I'd feel more comfortable if the exam was on paper, not online.					
4. I find it hard to invigilate/do relevant tasks when doing an online exam.					
5. I'd rather do exams on a computer than on paper, because I am used to working online.					
6. Online assessment is appropriate for my subject area which is English.					
7. My subject area/ English is too complex to be dealt with by online multiple-choice questions.					
8. Online exams don't just test knowledge of the subject, but IT skills as well.					

9. Online exams facilitate more authentic assessment than traditional methods through integration of multimedia, simulations, etc.					
10. Because students can guess the answer, online multiple-choice questions don't really reflect their level of knowledge.					
11. Online assessments use less paper, which is important to me.					
12. Technical problems make online exams impractical.					
13. Computer/Zoom/Internet fatigue makes online assessments impractical.					
14. It isn't practical for students to do online exams in the computer labs/ dormitory rooms/ libraries.					
15. Online exams are more practical than paper based exams because they are free from time and space.					
16. Marking is more accurate, because computers don't suffer from human error.					
17. The technology used in online assessments is unreliable.					
18. Online assessments favour some students more than others.					
19. Paper-based exams are fairer than online exams.					
20. Randomised questions from a bank means that sometimes students get easier questions.					
21. The test materials and results of online assessment are just as secure as paper-based assessment.					
22. The technology used in online exams is sufficiently effective in dealing with cheating and plagiarism.					
23. It is easier to cheat on online exams than with paper-based exams.					

24. The online exam system is vulnerable to hackers.					
25. Username and password login provide adequate security for online exams.					
26. The potential for immediate feedback with online assessment could help students learn.					
27. Online assessment facilitates a more adaptive learning approach than paper-based exams.					
28. Online assessment can add value to students' language learning.					
29. Online assessment is just a gimmick that does not really benefit learning and teaching.					
30. Online assessment goes hand-in-hand with e-learning (eg, using Blackboard/ Zoom/ Moodle etc).					

1: Strongly Disagree, 2: Disagree, 3: Neutral, 4: Agree, 5: Strongly Agree

APPENDIX B.

Approval Form Received From Istanbul Medeniyet University Educational Sciences Ethics Committee

İSTANBUL MEDENİYET ÜNİVERSİTESİ	**T.C.** **İSTANBUL MEDENİYET ÜNİVERSİTESİ** **EĞİTİM BİLİMLERİ ETİK KURULU KARARI**

Tarihi:	**Sayısı**
/ /	/

ARAŞTIRMA BAŞLIĞI	**Online Testing and Assessment in the English as a Foreign Language Context: Teachers' Perspectives** (*Yabancı Dil Olarak İngilizce Öğretiminde Çevrimiçi Ölçme ve Değerlendirme: Öğretmen Perspektifi*)"
ARAŞTIRMANIN TÜRÜ	Nicel araştırma (ilişki arayıcı)
ARAŞTIRMACI(LAR)	**İrem Gedil**
KARAR	

Görevi	**Kurul Üyeleri**	**İmza**
Başkan	Prof. Dr. Halil İbrahim SAĞLAM	
Üye	Prof. Dr. Ahmet AKIN	
Üye	Prof. Dr. Özlem FEDAİ	
Üye	Prof. Dr. Selami AYDIN	
Üye	Prof. Dr. Yeşim GÜLEÇ ASLAN	
Üye	Doç. Dr. Özge CENGİZ	
Üye	Doç. Dr. Özcan Erkan AKGÜN	
Üye	Av. Fevziye DOĞRUER	

APPENDIX C.

Descriptive statistics for the scale items (N=302)

Items		Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Mean	Standard Deviation
1. Using a computer adds to the stress of exams for teachers.	N	24	63	64	125	26	3.21	1.11
	%	7.9	20.9	21.2	41.4	8.6		
2. I expect computers to be used as part of regular assessment at university.	N	15	48	81	116	42	3.40	1.06
	%	5	15.9	26.6	38.4	13.9		
3. I'd feel more comfortable if the exam was on paper, not online.	N	9	32	59	110	92	3.80	1.07
	%	3	10.6	19.5	36.4	30.5		
4. I find it hard to invigilate/do relevant tasks when doing an online exam.	N	24	68	57	122	31	3.22	1.14
	%	7.9	22.5	18.9	40.4	10.3		
5. I'd rather do exams on a computer than on paper, because I am used to working online.	N	57	138	59	36	12	2.36	1.04
	%	18.9	45.7	19.5	11.9	4		
6. Online assessment is appropriate for my subject area which is English.	N	35	83	74	91	19	2.92	1.13
	%	11.6	27.5	24.5	30.1	6.3		
7. My subject area/ English is too complex to be dealt with by online multiple-choice questions.	N	23	88	43	96	52	3.21	1.24
	%	7.6	29.1	14.2	31.8	17.2		
8. Online exams don't just test knowledge of the subject, but IT skills as well.	N	10	61	60	138	33	3.40	1.03
	%	3.3	20.2	19.9	45.7	10.9		

Items		Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Mean	Standard Deviation
9. Online exams facilitate more authentic assessment than traditional methods through integration of multimedia, simulations, etc.	N	16	78	84	114	10	3.07	.98
	%	5.3	25.8	27.8	37.7	3.3		
10. Because they can guess the answer, online multiple-choice questions don't really reflect students' level of knowledge.	N	10	83	57	102	50	3.32	1.14
	%	3.3	27.5	18.9	33.8	26.6		
11. Online assessments use less paper, which is important to me.	N	7	34	56	137	68	3.74	1.00
	%	2.3	11.3	18.9	45.4	22.5		
12. Technical problems make online exams impractical.	N	3	30	34	143	92	3.96	.95
	%	1	9.9	11.3	47.4	30.5		
13. Computer/Zoom/Internet fatigue makes online assessments impractical.	N	6	56	53	125	62	3.59	1.06
	%	2	18.5	17.5	41.4	20.5		
14. It isn't practical for students to do online exams in the computer labs/ dormitory rooms/ libraries.	N	12	57	52	123	58	3.52	1.11
	%	4	18.9	17.2	40.7	19.2		
15. Online exams are more practical than paper based exams because they are free from time and space.	N	31	91	66	89	25	2.95	1.15
	%	10.3	30.1	21.9	29.5	8.3		
16. Marking is more accurate, because computers don't suffer from human error.	N	9	45	71	137	40	3.50	.99
	%	3	14.9	23.5	45.4	13.2		

Items		Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Mean	Standard Deviation
17. The technology used in online assessments is unreliable.	N	19	71	77	110	25	2.83	1.07
	%	6.3	23.5	25.5	36.4	8.3		
18. Online assessments favour some students more than others.	N	28	76	51	116	31	3.15	1.18
	%	9.3	25.2	16.9	38.4	10.3		
19. Paper-based exams are fairer than online exams.	N	10	67	59	108	58	3.45	1.13
	%	3.3	22.2	19.5	35.8	19.2		
20. Randomised questions from a bank means that sometimes students get easier questions.	N	10	64	86	126	16	3.24	.95
	%	3.3	21.2	21.2	41.7	5.3		
21. The test materials and results of online assessment are just as secure as paper-based assessment.	N	28	101	85	76	12	2.81	1.04
	%	9.3	33.4	28.1	25.2	4		
22. The technology used in online exams is sufficiently effective in dealing with cheating and plagiarism.	N	80	120	51	44	7	2.26	1.07
	%	26.5	39.7	16.9	14.6	2.3		
23. It is easier to cheat on online exams than with paper-based exams.	N	1	18	25	113	145	4.26	.87
	%	.3	6	8.3	37.4	48		
24. The online exam system is vulnerable to hackers.	N	6	25	66	146	59	3.75	.93
	%	2	8.3	21.9	48.3	19.5		
25. Username and password login provide adequate security for online exams.	N	34	122	92	59	4	2.56	.94
	%	11.3	40.4	30.5	16.6	1.3		

Items		Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Mean	Standard Deviation
26. The potential for immediate feedback with online assessment could help students learn.	N	2	33	66	166	35	3.65	.84
	%	.7	10.9	21.9	55	11.6		
27. Online assessment facilitates a more adaptive learning approach than paper-based exams.	N	7	73	109	98	15	3.13	.91
	%	2.3	24.2	36.1	32.5	5		
28. Online assessment can add value to students' language learning.	N	7	78	101	102	14	3.12	.92
	%	2.3	25.8	33.4	33.8	4.6		
29. Online assessment is just a gimmick that does not really benefit learning and teaching.	N	19	131	76	65	11	2.72	.98
	%	6.3	43.4	25.2	21.5	3.6		
30. Online assessment goes hand-in-hand with e-learning (eg, using Blackboard/ Zoom/ Moodle etc).	N	5	23	58	172	44	3.75	.85
	%	1.7	7.6	19.2	57	14.6		

APPENDIX D.

Tukey's Post Hoc Analysis for Items Indicating Statistically Significant Difference in Age

	Items	(I) Age Interval	(J) Age Interval	Mean Difference	Std. Error	Sig. (p-value)
Affective Factors	4. I find it hard to invigilate/do relevant tasks when doing an online exam.	20-30	31-40	.06	.20	.05
			41-50	.30	.22	
			*50+	.51	.24	
		31-40	20-30	-.06	.20	
			41-50	.24	.15	
			*50+	.44	.18	
		41-50	20-30	-.30	.22	
			31-40	-.24	.15	
			50+	.20	.20	
		50+	*20-30	-.51	.24	
			*31-40	-.44	.18	
			41-50	-.20	.20	
	5. I'd rather do exams on a computer than on paper, because I'm used to working online.	20-30	31-40	-.07	.20	.05
			41-50	.06	.22	
			50+	.39	.24	
		31-40	20-30	.07	.20	
			41-50	.13	.15	
			*50+	.46	.18	
		41-50	20-30	-.06	.22	
			31-40	-.13	.15	
			50+	.33	.20	
		50+	20-30	-.39	.24	

			*31-40	-.46	.18	
			41-50	-.33	.20	
Validity	6. Online assessment is appropriate for my subject area which is English.	20-30	31-40	.01	.20	.00
			41-50	.31	.22	
			*50+	.59	.24	
		31-40	20-30	-.01	.20	
			41-50	.30	.15	
			*50+	.57	.18	
		41-50	20-30	-.31	.22	
			31-40	-.30	.15	
			50+	.27	.20	
		50+	*20-30	-.59	.24	
			*31-40	-.57	.18	
			41-50	-.27	.20	
Practicality	11. Online assessments use less paper, which is important to me.	20-30	31-40	-.02	.18	.00
			41-50	.22	.19	
			*50+	.56	.21	
		31-40	20-30	.02	.18	
			41-50	.25	.14	
			*50+	.58	.16	
		41-50	20-30	-.22	.19	
			31-40	-.25	.14	
			50+	.33	.17	
		50+	*20-30	-.56	.21	
			*31-40	-.58	.16	
			41-50	-.33	.17	
		20-30	31-40	.06	.15	

Affective Factors	26. The potential for immediate feedback with online assessment could help students learn.		41-50	.19	.16	.00
			*50+	.54	.18	
		31-40	20-30	-.06	.15	
			41-50	.13	.11	
			*50+	.47	.13	
		41-50	20-30	-.19	.16	
			31-40	-.13	.11	
			50+	.34	.14	
		50+	*20-30	-.54	.18	
			*31-40	-.47	.13	
			41-50	-.34	.14	
	28. Online assessment can add value to students' language learning.	20-30	31-40	-.20	.17	.01
			41-50	-.00	.18	
			50+	.29	.19	
		31-40	20-30	.20	.17	
			41-50	.19	.13	
			*50+	.49	.14	
		41-50	20-30	.00	.18	
			31-40	-.19	.13	
			50+	.30	.16	
		50+	20-30	-.29	.19	
			*31-40	-.49	.14	
			41-50	-.30	.16	

APPENDIX E.

Tukey's Post Hoc Analysis for Items Indicating Statistically Significant Difference in Teaching Experience

	Items	(I) Teaching Experience Interval	(J)Teaching Experience Interval	Mean Difference	Std. Error	Sig. (p-value)
Affective Factors	1. Using a computer adds to the stress of exams for teachers.	<10	11-20	.28	.20	.00
			21-30	.19	.22	
			*>30	.82	.24	
		11-20	<10	-.28	.20	
			21-30	-.09	.15	
			*>30	.53	.18	
		21-30	<10	-.19	.22	
			11-20	.09	.15	
			*>30	.63	.20	
		>30	*<10	-.82	.20	
			*11-20	-.53	.22	
			*21-30	-.63	.24	
	3. I'd feel more comfortable if the exam was on paper, not online.	<10	11-20	.13	.16	.00
			21-30	.07	.18	
			*>30	.79	.24	
		11-20	<10	-.13	.16	
			21-30	-.06	.16	
			*>30	.65	.23	
		21-30	<10	-.07	.18	
			11-20	.06	.16	
			*>30	.71	.24	
		>30	*<10	-.79	.24	
			*11-20	-.65	.23	
			*21-30	-.71	.24	

	4. I find it hard to invigilate/ do relevant tasks when doing and online exam.	<10	11-20	.20	.16	.00
			21-30	.42	.18	
			*>30	.95	.24	
		11-20	<10	-.20	.16	
			21-30	.21	.16	
			*>30	.74	.23	
		21-30	<10	-.42	.18	
			11-20	-.21	.16	
			*>30	-.53	.24	
		>30	*<10	-.95	.24	
			*11-20	-.74	.23	
			*21-30	-.53	.24	
Practicality	11. Online assessments use less paper, which is important to me.	<10	11-20	.06	.14	.01
			21-30	.31	.16	
			*>30	.63	.21	
		11-20	<10	-.06	.14	
			21-30	25	.14	
			*>30	.55	.20	
		21-30	<10	-.31	.16	
			11-20	-.25	.14	
			>30	.31	.21	
		>30	*<10	-.63	.21	
			*11-20	-.56	.20	
			21-30	-.31	.21	
Impact on Teaching and Learning	28. Online assessment can add value to students'	<10	11-20	.12	.13	.03
			21-30	.28	.15	
			*>30	.54	.20	
		11-20	<10	-.12	.13	
			21-30	16	.13	
			>30	.41	.18	

	language learning.	21-30	<10	-.28	.15
			11-20	-.16	.13
			>30	.25	.20
		>30	*<10	-.54	.13
			11-20	-.41	.15
			21-30	-.25	.20

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı, Soyadı: İrem Gedil

Uyruğu: TC

Doğum Tarihi:

e-posta:

EĞİTİM

Derece	Kurum	Mezuniyet Yılı
Lisans	Orta Doğu Teknik Üniversitesi, Eğitim Fakültesi, İngiliz Dili Eğitimi Bölümü	2008
Yüksek Lisans	İstanbul Medeniyet Üniversitesi, Lisansüstü Eğitim Enstitüsü, İngiliz Dili Eğitimi Bölümü	Devam Ediyor

İŞ TECRÜBESİ

Tarih	Kurum	Görev
2019-	Sabancı Üniversitesi- Diller Okulu	Öğretim Görevlisi
2010-2019	İstanbul Şehir Üniversitesi- Diller Okulu	Öğretim Görevlisi
2008-2010	Bilkent Üniversitesi- Diller Okulu	Öğretim Görevlisi

Yabancı Diller

İleri Düzeyde İngilizce, Başlangıç Düzeyinde Almanca ve Fransızca