

ASSESSMENT LITERACY PERCEPTIONS OF TERTIARY LEVEL
ENGLISH LANGUAGE PROFESSIONALS

PINAR KILIÇ

A THESIS SUBMITTED
FOR
THE DEGREE OF MASTER OF ARTS
IN
CURRICULUM AND INSTRUCTION

İHSAN DOĞRAMACI BILKENT UNIVERSITY

ANKARA

SEPTEMBER 2022



To my beloved family...

Assessment Literacy Perceptions of Tertiary Level English Language Professionals

The Graduate School of Education
of
İhsan Doğramacı Bilkent University

by

Pınar Kılıç

In Partial Fulfilment of the Requirements for the Degree of
Master of Arts
in
Curriculum and Instruction
Ankara

September 2022

İHSAN DOĞRAMACI BILKENT UNIVERSITY
GRADUATE SCHOOL OF EDUCATION

Thesis Title: Assessment Literacy Perceptions of Tertiary Level English Language

Professionals

Pınar Kılıç

September 2022

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Curriculum and Instruction.

Asst. Prof. Dr. Tijen Akşit (Advisor)

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Curriculum and Instruction.

Assoc. Prof. Dr. Perihan Savaş, Middle East Technical University (Examining Committee Member)

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Curriculum and Instruction.

Asst. Prof. Dr. Donald Staub, İzmir University of Economics (Examining Committee Member)

Approval of the Graduate School of Education

Prof. Dr. Orhan Arıkan (Director)

ABSTRACT

ASSESSMENT LITERACY PERCEPTIONS OF TERTIARY LEVEL ENGLISH LANGUAGE PROFESSIONALS

Pınar Kılıç

M.A. in Curriculum and Instruction

Advisor: Asst. Prof. Dr. Tijen Akşit

September 2022

The aim of this study was to explore the assessment literacy perceptions of English language professionals. This mixed method case study was conducted through surveys and interviews with English language professionals at the English preparatory school of a foundation university in Ankara, Turkey. The survey items were adapted from Kremmel and Harding (2020) with the purpose to explore the perceived assessment literacy levels of English language teachers. The survey was completed by 23 English language teachers, four assessment specialists and seven academic leaders. In addition, 10 teachers, four assessment specialists and seven academic leaders were interviewed to explore the perceived relationship between their assessment literacy and teaching quality. Descriptive statistics were used for quantitative data analysis. The interview data were categorized based on the assessment literacy dimensions specified in Taylor's framework (2013) and analyzed through thematic content analysis. The results indicated that English language teachers are moderately knowledgeable and slightly skilled in assessment literacy dimensions. The study also concluded that English language professionals establish a meaningful relationship between several dimensions of assessment literacy and their teaching quality. The overall results suggest that English language professionals need to be provided with further training to improve their assessment literacy and to empower assessment for learning.

Keywords: Assessment literacy, English language assessment, assessment for learning

ÖZET

YÜKSEK ÖĞRETİM İNGİLİZ DİLİ UZMANLARININ ÖLÇME DEĞERLENDİRME OKURYAZARLIK ALGILARI

Pınar Kılıç

Eğitim Programları ve Öğretim Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi Tijen Akşit

Eylül, 2022

Bu çalışmanın amacı, yüksek öğretim İngilizce dili uzmanlarının değerlendirme okuryazarlığı algılarını araştırmaktır. Bu karma yöntemli vaka çalışması, Ankara, Türkiye'de bir vakıf üniversitesinin İngilizce hazırlık okulundaki İngiliz dili uzmanlarıyla anketler ve görüşmeler yoluyla gerçekleştirildi. Anket maddeleri, İngilizce öğretim görevlilerinin, algılanan ölçme değerlendirme okuryazarlık düzeylerini araştırmak amacıyla Kremmel ve Harding (2020)'den uyarlandı. Anket 23 İngilizce öğretim görevlisi, dört ölçme değerlendirme uzmanı ve yedi akademik lider tarafından tamamlandı. Ayrıca, ölçme değerlendirme okuryazarlığı ile öğretim kalitesi arasındaki algılanan ilişkiyi araştırmak için 10 İngilizce öğretim görevlisi, dört ölçme değerlendirme uzmanı ve yedi akademik liderle görüşmeler yapıldı. Nicel veri analizi için tanımlayıcı istatistikler kullanıldı. Görüşme verileri Taylor (2013) çerçevesinde belirlenen ölçme değerlendirme okuryazarlığı boyutlarına göre kategorize edildi ve tematik içerik analizi yöntemi ile analiz edildi. Sonuçlar, İngilizce öğretim görevlilerinin ölçme değerlendirme okuryazarlığı boyutlarında orta düzeyde bilgili ve biraz yetenekli olduğunu gösterdi. Görüşmelerin bulguları, İngiliz dili uzmanlarının ölçme değerlendirme okuryazarlığının çeşitli boyutları ile öğretim kaliteleri arasında anlamlı bir ilişki kurduğunu ortaya koydu. Çalışmanın genel sonuçları aynı zamanda İngiliz dili uzmanlarının ölçme değerlendirme okuryazarlıklarını geliştirmek ve öğrenme için ölçme değerlendirmeyi güçlendirmek için daha fazla eğitim verilmesi gerektiği sonucuna vardı.

Anahtar kelimeler: Ölçme değerlendirme okuryazarlığı, İngilizce ölçme

değerlendirme, öğrenme için ölçme değerlendirme

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Asst. Prof. Dr. Tijen Akşit for the continuous support, guidance, and immense knowledge she provided throughout my research process. It has been a privilege to be her student, and this thesis would not be as well-founded without her expertise.

Secondly, I thankfully acknowledge the feedback and insights provided by the examining committee members, Assoc. Prof. Dr. Perihan Savaş and Asst. Prof. Dr. Donald Staub, to make my thesis more forceful. In addition, I want to thank my director, Dr. Taner Yapar, for his care and motivation all throughout my career and education.

I would also like to thank my dear family starting with my mother Akile Kılıç, a retired English language teacher who devoted much of her energy to her students for years. She has been a great inspiration to build up my career as my first-ever English teacher. Besides, my father Mustafa Kılıç has always been there whenever I needed to be encouraged and heartened, so I owe him a big thanks. My dear sisters Zeynep Derya Tepeli and Ayşenur Kılıç, who are also very successful and ambitious English language teachers, have always been a source of thankfulness and appreciation all throughout my life. Last but not least, my deepest love goes to my fiancée Gürcan Ateş, who makes the world a happier place for me and always promotes my personal and professional development.

As friends are the family we choose, my sincere thanks also go to my long-term friends for their support, care, and firm confidence in me. Moreover, my officemates deserve much acknowledgment for their valuable contribution and feedback on my research.

Finally, I would like to thank each and every entity in this universe that has shaped me into who I am today.

TABLE OF CONTENTS

ABSTRACT	iii
ÖZET	IV
ACKNOWLEDGEMENTS	V
TABLE OF CONTENTS	VI
LIST OF TABLES	XI
LIST OF FIGURES	XIII
CHAPTER 1: INTRODUCTION	1
Introduction.....	1
Background	2
Problem	4
Purpose.....	5
Research Questions	5
Significance.....	6
Definition of Key Terms	7
CHAPTER 2: REVIEW OF RELATED LITERATURE.....	8
Introduction.....	8
Assessment in Curriculum	10
Assessment of Learning: Curriculum as Product	10
Assessment for Learning: Curriculum as Process	11
Assessment as Learning: Curriculum as Praxis.....	12
Washback.....	12
Language Assessment	13
Classroom Assessment	15
Formative Assessment.....	16

Summative Assessment	18
Alternative Assessment/Portfolio Assessment	19
Integrated Assessment	22
Assessment of Language Proficiency	24
Assessment Literacy	26
Theoretical Framework of the Study	28
Assessment Literacy for Teachers' Development	31
Perceptions of Assessment Literacy	33
Conclusion	36
CHAPTER 3: METHOD	38
Introduction.....	38
Research Design.....	38
Context.....	41
Assessment Regulations	41
Assessment Schedule.....	42
The Unique Setting.....	43
Participants.....	45
Instrumentation	49
Surveys	50
Interviews	53
Method of Data Collection.....	54
Methods of Data Analysis.....	55
CHAPTER 4: RESULTS	58
Introduction.....	58
Perceived Assessment Literacy Level of English Language Teachers.....	59

Survey results	60
Interview Results	72
Knowledge of Theory	74
Language Pedagogy	75
Principles and Concepts.....	76
Impact of Social Values	77
Local Practices	78
Personal Beliefs and Attitudes	78
Scores and Decision Making	80
Constructing Language Assessments.....	80
Administering and Scoring Language Assessments	81
Evaluating Language Assessments	81
English Language Professionals' Perception of the Relationship Between Assessment Literacy and Teaching Quality.....	82
English Language Teachers' Perception	82
Knowledge of Theory	82
Language Pedagogy	83
Principles and Concepts.....	88
Impact of Social Values	91
Local Practices	91
Personal Beliefs and Attitudes	92
Scores and Decision Making	93

Constructing Language Assessments.....	94
Administering and Scoring Language Assessments.....	95
Evaluating Language Assessments.....	95
Assessment Specialists' Perceptions	96
Perceived Assessment Literacy Level of Assessment Specialists	96
Assessment Specialists' Perception of the Relationship Between Assessment Literacy and Teaching Quality.....	102
Academic Leaders' Perception.....	109
Perceived Assessment Literacy Level of Academic Leaders	109
Academic Leaders' Perception of the Relationship Between Assessment Literacy and Teaching Quality	118
Summary of the Findings.....	132
CHAPTER 5: DISCUSSION.....	141
Introduction.....	141
Overview of the Study	141
Discussion of Major Findings.....	141
English Language Teachers' Assessment Literacy Level	141
The Relationship Between Assessment Literacy and Teaching Quality .	144
Implications for Practice	149
Implications for Further Research	151
Limitations of the Study.....	152
REFERENCES.....	153
APPENDICES	170
Appendix A.....	170

Appendix B	178
Appendix C	186
Appendix D	193
Appendix E	195



LIST OF TABLES

Table	Page
1 Factors Influencing Teacher Assessment Literacy in Questioning Practice.....	32
2 Demographic Information of English Language Professionals (Survey)...	46
3 Background Information of English Language Professionals (Survey).....	48
4 Demographic Data of English Language Professionals (Interview).....	49
5 Number of Survey Items per Each Assessment Literacy Dimension.....	50
6 Cronbach's Alpha Values for Each Assessment Literacy Dimension in the Survey	52
7 Matching of Research Questions with the Data Sources	59
8 Assessment Literacy Levels of English Language Teachers.....	60
9 Means and Standard Deviations for Knowledge and Skills Categories.....	61
10 Descriptive Statistics for the Dimensions in the Assessment Literacy Survey.....	63
11 Means and Standard Deviations for Assessment Literacy Levels: Knowledge Items.....	65
12 Means and Standard Deviations for Assessment Literacy Levels: Skills Items.....	69
13 Summary of Perceived Assessment Literacy Levels Based on Dimensions.....	136
14 References to the Assessment Literacy Dimensions in relation to Teaching Quality.....	138

15 Most Frequently Covered Concepts of Assessment Literacy

Dimensions.....	139
-----------------	-----



LIST OF FIGURES

Figure	Page
1 Differential Language Assessment Literacy Profile.....	29
2 Decision Impact of Assessment-Literate Educators.....	36
3 Convergent Mixed Method Procedures	39
4 Sample Analysis of Interview Data.....	56
5 Comparison of the Results with the Differentials in Taylor's Framework (2013)	132
6 Assessment Specialists' Profile in Taylor's Framework (2013)	134
7 Academic Leaders' Profile in Taylor's Framework (2013)	135

CHAPTER 1: INTRODUCTION

Introduction

One needs to understand and acknowledge inevitable preferences in order to know their influence so that fair treatment of data and arguments can be attained.

Stephen J. Gould, 1981, p.36

Throughout the world, certain types of data are used merely to rank or place students. In this way, students come to fear assessments as a judgment or verdict of their learning achievements. This application of assessment is destructive and often discourages students from learning. Many English teachers fail to reflect their intended course principles in their assessment practices due to various reasons including lack of assessment literacy (Earl, 2013). Over the past decades, researchers have come up with convincing evidence that suggests that teaching and learning could be empowered with the help of assessments when employed appropriately to monitor student learning and make the necessary adjustments accordingly (Popham, 2018). Moreover, the responsibility attained to language teachers in the assessment processes has increased as a result of the expanded use of language assessments at an international scope (Fulcher, 2012). However, relevant literature suggests that a large number of English language teachers do not have comprehensive understanding of the fundamental concepts of assessment literacy, and how assessment can be used as a teaching strategy to facilitate student learning (Coombe et al., 2020).

In the era of knowledge and development, it is of vital importance for teachers to be aware of their assessment literacy experiences. Assessment literate teachers and teacher trainers maximize autonomy, motivation and student achievement (Stiggins, 2002). Newfields (2006) lists three reasons why assessment literacy matters for teachers. Firstly, assessment practices are worth a lot of time and money for most of the educational contexts, so teachers need to understand the rationale behind assessment decisions. Secondly, for the educational literature to be better interpreted, teachers should be capable of evaluating the statistical concepts in the right way. A final reason listed takes us to a 21st-century skill: Communication. Newfields argues that assessment literacy allows teachers to communicate assessment results to a wider audience and make self-critique (2006). Thus, a growing body of literature has focused on how teachers can improve their assessment literacy through pre-service and in-service training, assessment courses, self-study and so on so forth (Kremmel & Harding, 2020).

Background

Assessment in education refers to “the process of gathering evidence of student achievement to inform educational decisions” (Stiggins, 1991). For many years, unfortunately, assessment has been viewed as a means to grade students to label them according to their test scores. Over the past few decades, concepts of assessment have begun to change, especially with the widespread use of formative assessment tools (Sevimel-Sahin, 2021). Especially, language teaching and language assessment has gained further importance and consideration with the changing dynamics of the world such as increasing number of people studying in English medium institutions and immigrating to other countries for work (Kunnan, 2012). Therefore, the position of English language as the major international language has

increased the number of language assessments conducted globally (Spolsky, 2012). The widespread use of English language and assessments at tertiary levels led the public and private organizations such as universities to offer comprehensive language programs for individuals (Stanevich, 2008). This effect is also visible in Turkish context where the number of English language courses is increasing, and many universities require their students to study English preparatory classes to be eligible to continue to their majors (Öz & Atay, 2017). In this regard, teachers have critical roles since they are the ones who can refer to the relevant knowledge to help learners improve with assessment (Shams & Ikbāl, 2019).

As Stiggins (2014) claims, assessment should be used not only to determine student success, but also to increase motivation and improve the quality of education. This claim follows a notion Stiggins (1991) made when he introduced the term assessment literacy. Popham (2018) further explains assessment literacy as an individual's understanding of the fundamental assessment concepts and procedures deemed likely to influence educational decisions. Assessment literacy has been explored for various purposes ranging from instructional improvement to professional development. The concept provides teachers with the potential to benefit from developing and implementing language assessment literacy in their classroom practice; these practices could help define students' needs, increase their motivation and provide feedback for achievement (Tsagari, 2021). It has gained attention as more effort has been put forth to use assessment for instructional purposes.

With growing awareness of the social responsibilities of the language assessment profession, this has led to calls for professionals to work together to improve the assessment literacy of teachers, policy-makers and others and

to provide better public information on assessment issues. (Green, 2021, p. 232)

Language assessment literacy is a complicated construct to conceptualize for all stakeholders (Giraldo, 2020). Relevant literature shows that some EFL teachers are not much insightful about the concept of assessment literacy (Sevimel-Sahin, 2021). Insufficient assessment literacy among teachers and their long-held beliefs that only specialists, not teachers, are responsible are assessment results in missed opportunities for using assessment for learning (Herrera Mosquera & Macías, 2015).

Problem

Even though language teaching is a long-time profession, language assessment is relatively a young discipline (Fulcher & Davidson, 2012). Relevant literature shows that English language teachers are not provided with enough pre-service and in-service training on assessment (Taylor, 2009; Turner, 2012). Therefore, many English language teachers lack the necessary assessment literacy skills to benefit their teaching at practical levels. Furthermore, measurement and evaluation are often seen as ways to determine passing and failing students, which is a process often conducted by a separate unit or group of test developers (Popham, 2018). The lack of training in assessment and the idea that assessment is a completely different and separate practice from teaching inhibit language teachers from making use of assessments as a learning tool in their classes. Assessment does not attract its well-worth attention at the level of its use for increasing the efficiency of teachers' and learners' performances (Tsayari & Vogt, 2017). However, assessment does more than allocating numerical values to students' achievement, it provides feedback on how much the institution reaches its course objectives (Aksit et al., 2018). For the overall amendment of quality teaching, teachers should improve

their understanding of assessment literacy and assessment for/as learning (Earl, 2013). Otherwise, it may restrict valuable feedback that could be used to advance teachers' professional development; in particular, this feedback could be used to develop teachers' assessment literacy. In addition to teachers, other stakeholders such as test developers and educational leaders require expanded knowledge and skills to support assessment for learning (Taylor, 2013). Although recommendations for the use of language assessment literacy for teaching practices can be found in literature (Guo & Xu, 2021; Hamp-Lyons, 2017), it is still an area that needs more consideration and improvement (Giraldo, 2020).

Purpose

The main purpose of this case study was to explore the assessment literacy perceptions of English language professionals at a preparatory school of a foundation university in Ankara. Since in-depth data would be useful to investigate the perceptions, a mixed-method study was designed. Through surveys and semi-structured interviews, English language teachers', assessment specialists' and academic leaders' perceptions on their assessment literacy levels and its relationship to their teaching quality have been explored.

Research Questions

This study addresses the following research questions:

1. What is the assessment literacy level of English language teachers as perceived by
 - a. themselves,
 - b. assessment specialists, and
 - c. academic leadersat the language preparatory school of a foundation university in Ankara?

2. How do

- a. English language teachers,
- b. assessment specialists, and
- c. academic leaders

perceive the relationship between assessment literacy and teaching quality?

Significance

The English language is more than a subject that is taught in schools; it has become essential to understand the readings or course materials in various contexts, ranging from K-12 to higher education. In the globalizing world, many students also choose to study at English-medium schools and universities where they are taught all academic subjects in the English language. In this respect, foreign language assessment, specifically English language assessment has gained more consideration over the recent years. Therefore, a study on language assessment literacy of tertiary level English language professionals would provide insight for English as a foreign language teachers, assessment specialists and academic leaders working at tertiary level both in Turkey and around the world; and help increase awareness of assessment issues.

The exploration of assessment literacy perceptions from the perspectives of teachers, assessment specialists and academic leaders could yield valuable information for all stakeholders so as to foster assessment for and assessment as learning. Establishing an understanding of the term assessment literacy and owning it for practical uses in higher education will help institutions reach their course objectives more smoothly (Stiggins, 2014). Almost all teachers have assessment related job descriptions such as giving feedback to students after certain examinations or other assessment practices. In addition, many include in-class

assessment strategies in their lesson plans for various purposes including student motivation. Raising their awareness and knowledge on the issue of assessment practices could help the teachers become more self-assured. With more intuition about assessment literacy, teachers could better interpret the assessment results, modify instruction based on assessment results, and thus increase student achievement (Coombe et al., 2020). The findings of the study could provide insightful recommendations and practical solutions for all stakeholders in order to empower teaching and learning through assessment literacy in their job practices. The findings of this study could potentially encourage the English preparatory schools at tertiary level institutions to integrate the concept of assessment literacy into their in-service teacher training programs effectively. Therefore, the results of this study could help both define EFL teachers' assessment literacy levels and provide them with the insight to better apply assessment for learning principles to better respond to the needs of the learners and institutions (Herrera & Macías, 2015).

Definition of Key Terms

Assessment: The process of gathering evidence of student achievement to inform educational decisions (Stiggins, 1991).

Assessment literacy: An individual's understanding of the fundamental assessment concepts and procedures deemed likely to influence educational decisions (Popham, 2018).

Teaching quality: Instruction that leads to effective learning, which in turn means thorough and lasting acquisition of the knowledge, skills, and values the instructor or the institution has set out to impart (Felder & Brent, 1999).

CHAPTER 2: REVIEW OF RELATED LITERATURE

Introduction

In this chapter, a review of related literature is presented in addition to research findings and discussions relevant to the concepts in this study. Starting with a brief introduction of the fundamentals of assessment, the chapter explores the place of assessment in curriculum. Following the comparison of three major approaches to assessment in curriculum, this chapter covers several assessment methods including classroom assessment, formative assessment, summative assessment, alternative assessment, integrated assessment and assessment of language proficiency. In addition, the concept of assessment literacy and its relationship to quality language teaching are presented and discussed with the relevant studies on assessment literacy perceptions in this chapter.

Assessment has been an essential part of educational settings for a long time for various purposes and in different forms (McNamara, 2000). It refers to the broadest extent of practices to define a student's competence, as defined by Popham (2018). Quality assessment practices fundamentally require to be valid and reliable, with an aim to have beneficial consequences for teaching (Green, 2021). Validity refers to the extent to which the assessment tools measure what they claim to measure (Fulcher & Davidson, 2007). It does not only encompass the technical considerations but also the relevance, and usefulness of the assessment results (Chapelle, 2012). Reliability, at its simplest form, means the consistency of test performance reflected in the assessments (Hughes, 1989). Educators aim to make sure that the assessment results are generalizable over the samples of assessment

tools, scorers and so forth for reliable assessment (Kane, 2012). The concept of reliability is also closely discussed with the reference points of assessments as norm-referenced assessment and criterion-referenced assessment. Norm-referenced assessment refers to interpreting a learner's performance in comparison to other learners, whereas criterion-referenced assessment involves applying to external criterion (Jones, 2012), such as a proficiency framework like the Common European Framework of Reference for Languages (CEFR).

Along with the changing dynamics of educational settings as well as school culture, governmental policies and so forth, assessment is also subject to continuous adaptation (Rea-Dickins, 2006). This adaptation is not only limited to curricular practices regarding students such as examinations, but also bears requirements for teachers applying the curriculum (Llosa, 2011). To promote teaching and learning through assessment, understanding the place of assessment in curriculum and how assessment literacy helps teachers to improve would provide valuable insights for curriculum development and teacher improvement as well as further research (Giraldo, 2020).

Along with the importance placed on assessment, the term assessment literacy has also gained consideration and is sure to have an impact on how successful assessment policies are conducted (Jiang, 2020). To promote active participation of teachers in the assessment procedures, language teachers must be aware of the principles of assessment literacy and effectively make use of it in their teaching practices (Stiggins, 2002). Green (2018) also mentions this need by referring to the responsibility of teachers to take a role in both assessing their learners in the classroom and also preparing them for assessment. Different kinds of assessment strategies are practiced by teachers, and each teacher deals with the

results of assessment in diverse ways. Therefore, assessment literacy must be integrated into continuous professional development activities more frequently (Coombe et al., 2020). Jiang lists the common factors impacting a teacher's assessment practices as stakeholders' needs, the school culture and professional development (2020). These factors are also effective on different approaches to curriculum (Earl, 2013). In other words, the role of students and teachers in assessment implementations as well as their motivation to benefit from assessment literacy are divergent in different curricula (Fulcher & Davidson, 2007).

Assessment in Curriculum

Assessment of Learning: Curriculum as Product

Assessment of learning is perhaps the oldest and most traditional way to look at assessment. As Davidson (2018) defines, assessment of learning is the summative practices conducted at the end of the learning processes. The purpose of assessment is often to label students with certain scores obtained from tests and examinations. These tools are utilized to determine whether students are capable of these measured skills and competencies. In this respect, assessment of learning worked well back in the past when curriculum as a product gained popularity with the rise of 'scientific management' in the mid-1900s (Smith, 1996). Curriculum as product involves systematic formulation of behavioral objectives broken into smaller units in order to be measured.

Assessment of learning and curriculum as product are both teacher centered. As criticized by Smith, in this type of approach, assessment may resemble a shopping list with skills to be ticked, mostly through test items, which are inevitable components of assessment of learning (1996). Assessment of learning has also received criticism in terms of incapability of providing teachers with meaningful data

for policy and decision making (Stiggins, 2002). Therefore, a more accountable and powerful approach to assessment and curriculum in general was definitely in demand.

Assessment for Learning: Curriculum as Process

Another approach to looking at curriculum is as a process and involves assessment for learning rather than assessment of learning. Curriculum as process sees curriculum not as a physical entity, but rather as the interaction of teachers, students and knowledge (Smith, 1996). Dependent on behavioral objectives, it aims to benefit from experimentation. This model is not as teacher-centered as curriculum as product; as a result, assessment practices also shift the focus away from the teacher, which can also be observed in assessment for learning.

Unlike assessment of learning, assessment for learning shifts the focus from summative to formative practices. Earl (2013) states that teachers refer to a different range of data to measure students' success and also make use of these data for further learning. The tools teachers use for assessment such as worksheets, student-teacher conferences, and so on so forth, provide feedback not only on what students know but also on how the rest of learning should be designed (Earl, 2013). It is interactive, not teacher-centered. This focus on student-teacher interaction and continuous feedback can also be observed in curriculum as process. Smith (1996) maintains that in the curriculum as process, teachers continually evaluate the process of their learners and what they can do with those outcomes.

To sum up, the aim of assessment for learning is not to make comparative judgments among students, but to clarify each student's understanding and thinking, which will provide valuable insights for their further learning (Earl, 2013).

Assessment as Learning: Curriculum as Praxis

The last approach to look at curriculum, curriculum as praxis, bears its focus on the ‘emancipation’ of the students as Smith (1996) mentions in his article. In curriculum as praxis, there is reciprocity in all of the components of such curriculum as planning, acting and evaluating. Teachers of this ideology continually evaluate their students and take a further step by contemplating on what they can do to improve learning through the evaluation outcomes.

Similarly, assessment as learning encourages students to have active roles in their own learning experiences where they are self-motivated and should take initiative in these experiences. Assessment as learning supports the development of meta-cognition and critical thinking by enabling the learners to reflect on their assessment as well as providing feedback for adjustments and adaptations as Earl (2013) suggests.

All three perspectives for curriculum worked well at certain times, cultures and contexts, and probably will go on doing so in the future. However, as Stiggins (2002) asserts, it is of high importance that new perspectives should be developed to link assessment and teaching quality in more purposeful ways. In this respect, assessment literacy will serve meaningful functions not only for student empowerment but also for teacher development.

Washback

Washback refers to the effects of assessment on teaching and learning (McNamara, 2000), mostly at the classroom level (Tsagari & Cheng, 2017). In a broader context, these effects can be observed within the school, institution, the education system and even the society. The washback effect is often considered negative, positive or neutral by the stakeholders in language teaching. Teachers

should be aware of these potential influences on their short-term and long-term practices in teaching (Green, 2006). When reflecting the course objectives and the teaching philosophies of the teachers teaching, washback can be considered a positive impact on teaching and learning. To illustrate, tests can yield results that inform teaching and learning processes (Atmojo, 2021). However, multiple choice types of tests are often found to restrict communicative approaches to language teaching and learning (McNamara, 2000). In this sense, the summative assessment does not aim to offer opportunities for teachers to shape their instruction accordingly (Ahmed et al., 2019). For instance, when the proficiency test is composed of only multiple-choice types of questions, teachers tend to ignore other skills such as writing or speaking. In many relevant studies, it was found that teachers believe their instruction was dominated by exam-focused materials and teaching (Tsagari, 2011). Furthermore, negative washback caused by proficiency tests leads the instructors to become a test teacher and only teach for tests (Alqahtani, 2021). Therefore, the literature suggests that formative assessments like portfolios have better washback than separate item formats (McNamara, 2000). One major reason for this is the nature of the formative assessments which require both the teacher and the student to allocate time *engaged* in the tasks. In addition, numerous strategies are suggested in the relevant literature such as improving teachers' assessment literacy (Hughes, 1989) and involving different stakeholders in the assessment processes (Shohamy, 1992), in order to promote positive washback.

Language Assessment

Language assessment has been a major global issue in the education world, which in turn created the need to redefine the relationship among assessment, curriculum, and teacher practice (Cumming, 2009). Due to the complexity of the

language construct, language assessment methods are fundamentally different from those in many other disciplines (Brown & Hudson, 1998). However, most language teaching programs aim to develop their learners' abilities in the 'traditional' four skills of listening, reading, writing and speaking (Green, 2021). Assessment of receptive skills and productive skills requires the definition of the knowledge and skills to be measured, identification of the language abilities that students should perform, and choosing the right tasks to enable learners to demonstrate their language skills (Green, 2021). As Bachman and Palmer (1996) asserted, language assessment should involve appropriate input to provide the learners such as the reading texts and audio recordings; and it should be authentic, reflecting the kind of language appearing in real life. In addition, assessment of productive skills should ensure that the tasks get the assesses to produce the language content so that their abilities could be reflected in the tasks (Green, 2021). However, assessment of productive skills, especially speaking skill has always been challenging for language professionals (Elder, 2017). Due to concerns related to practicality (Green, 2021), language assessment was predominantly in the form of multiple-choice type of items in the 1950s and 1960s; nevertheless, more communicative assessments like formative assessment and portfolios are being preferred by the teachers and policymakers in the current century (Vogt & Tsagari, 2014).

In many countries, there are standardized English tests that are internationally accepted for school admissions, job positions and so on so forth (Stanevich, 2008). For instance, the worldwide known company English Testing Service (ETS) has been conducting standardized tests such as the Test of English as a Foreign Language (TOEFL) for the last 70 years. Similarly, British Council conducts The International English Language Test (IELTS), which is a popular test for people who

would want to work or study in countries where English is the native language (Sawaki, 2012). As Elder (2017) highlights, there has been an increase in the number of universities requiring students to have minimum English entry requirements so that students are able to deal with their undergraduate studies better. In certain countries such as Turkey, this concept shows itself in the form of preparatory classes where students learn only English for specific periods of time, usually a year before they continue with their majors. As a consequence of the globally growing body of language programs, different models have been suggested for valid and reliable assessment of the English language (Chirimbu, 2013).

Classroom Assessment

With the shift of focus from the output to the process in English language assessment, teachers have had to be more involved in the planning and implementation of assessment processes in recent years (Akşit et al., 2018). Schools are now referring to their teachers to design classroom assessment strategies, especially as teams of their shared levels (Buckley-Walker & Lipscombe, 2021). Classroom assessment is different from large-scale assessment in the sense that in classroom assessment there is the teacher as a facilitator who engages with the learners along the learning and assessment processes. It is more authentic and meaningful than a high-stakes test because it can provide a wider and more informative reflection of student performance (Llosa, 2011). The personal knowledge and the learning process of the learners are taken into account so that assessment supplies future directions for learning. As teachers have the opportunity to observe their learners involve in varying activities, interactions and relationships in classrooms, they are able to collect as much evidence of learning as possible. Collaboration is a key component of classroom assessment where there is continuous

feedback, interaction and shared goals between teachers and learners. The evaluation is not conducted by the teacher solely, but by the learners and their peers as well; and rather than the traditional way of looking at the product, how the task is carried out is also a matter of subject in scoring (Fulcher & Davidson, 2007).

In their study with 746 Chinese teachers, Coombs et al. (2021) concluded that being actively involved in classroom assessment enabled teachers to stay up to date with the contemporary assessment methods and thus foster professional development. Fulcher and Davidson (2007) asserted that classroom assessment is directly relevant to a valid assessment of learners because the classroom itself is the context where learning happens and therefore progress can be best traced through experiences there. The authors describe two good characteristics of classroom assessment. Firstly, good classroom assessment designs are interlinked; that is, they build upon each other and the learning environment. Second, they should be implemented over a time period such as months or years. Besides, Wallace and Qin call for equally and fairly applied classroom assessment strategies with the use of courteous communication (2021). Rea-Dickins (2001) adds to these informal classroom assessment opportunities embedded in teachers' curriculum so as to detect how much a learner knows and how well s/he is progressing. And Buckley-Walker and Lipscombe (2021) stress the importance of alignment of classroom assessment to curriculum and instructional practices for stronger validity.

Formative Assessment

Formative assessment is usually defined as the assessment design which takes place during the learning process with the aim of encouraging further learning (Fulcher & Davidson, 2007). In formative assessment, both teachers and learners are aware of the learning intentions and criteria for success. While the teacher is

responsible for designing appropriate assessment tasks that would provide meaningful feedback to move learning forward, the learner is responsible for learning in that context, also as peers who cooperate for their own instruction (William & Thompson, 2017). Teachers can make use of feedback simultaneously while teaching or refer to students' homework results in order to plan the upcoming lesson accordingly (Black & Wiliam, 2009). Crisp (2011) also highlights that timely and meaningful feedback must be provided for students so as to have a positive impact on learning. This concurrence is the reason why formative assessment can also be called 'classroom assessment', which happens while the program is being implemented (He & Mei, 2020).

Although formative assessment is believed to serve learning purposes, a lot of research shows that it is not yet widely accepted and applied in EFL classes all around the world. In a study conducted with university teachers, Bashir, Karim and Akram (2020) worked with 40 prospective teachers and five experienced teachers by exploring their use of formative assessment through an observation schedule and a follow-up survey. The researchers concluded that the teachers were able to use formative assessment only at lower-level abilities of the cognitive domain of Bloom's Taxonomy; i.e. knowledge and comprehension. Similarly, Guo and Xu (2021) conducted a survey with Chinese instructors of English in writing classes at a university. The researchers investigated the formative assessment strategies used by 362 teachers focusing on how they provided feedback in writing classes, and how they perceived the aims of these assessments. The study concluded that very few teachers use formative assessment strategies such as journals and self-/peer assessment. Most teachers stated that they refer to students' writings as proofs of achievement, rather than see them as opportunities for improving learning. However,

it was also concluded that there is a positive correlation between the use of formative assessment and the teachers' previous participation in a language assessment course. Therefore, assessment literacy plays a crucial role in the effective implementation of formative assessment (Guo & Xu, 2021). This result was supported by Bergeron (2020) when the researcher carried out a mixed-method study with a survey and open-ended questionnaire. Teacher perceptions of formative assessment were explored, and the findings indicate that most teachers value the benefit of diagnosing learners' strengths and benefits through formative assessment. However, because the older teachers do not feel competent enough to make use of formative assessment strategies such as online quizzes to be used during class, they refrain from using them for effective purposes (Bergeron, 2020).

Summative Assessment

Summative assessment is the "assessment conducted at the end of a program of study to assess whether and how far individuals or groups have been successful" (Fulcher & Davidson, 2007). It is usually regarded as a proof and reflection of students' previous learning. Throughout history, it has been widely used, especially because it is a practical solution to classroom grading, and local, national and international assessment for various purposes (Stiggins, 2005). In addition, when the objectives are clearly defined, end-of-program tests can well inform the teachers about whether they have achieved these objectives and whether their instruction was effective (Connors, 2021). Related to the current status of learners, summative assessment enables teachers to evaluate learning against specific standards. In a cross-sectional study in the UK, 53 students were given summative assessment modules for seven weeks. Students used these assignments to prepare for their exams and the assignments made up 20% of their final grades. In the study, it was

concluded that summative assessment has a direct impact on students' self-efficacy and self-regulation in assessment tasks (Fergus et al., 2021). In another similar study with undergraduate students, learners were introduced continuous summative assessment as a module. During the application and right after, student perceptions were explored through questionnaires and interviews. Trotter (2006) concluded in this study that students were extremely positive about this summative assessment module and stated that it provided them with the necessary extrinsic motivation required in a learning environment.

Alternative Assessment/Portfolio Assessment

Alternative assessment is “a flexible assessment approach that is representative of the curriculum, meaningful to learners, and indicative of the learners' acquired ability or knowledge” (Holt, 1992, p. 135). Examples of alternative assessment in language classes include but are not limited to portfolios, presentations, mock interviews, reports, videos and projects (Rus, 2019). Other types of alternative assessment commonly mentioned in the literature are case studies, diaries and self-assessment, all of which have wider purposes and opportunities than traditional testing methods (Chirimbu, 2013). When Rus (2019) describes each example of alternative assessment in her article, she emphasizes that the common attempt is to assess students' language skills through authentic and communicative tasks. Different from traditional assessment of language where the focus is on how much the learner ‘knows’ about language domains such as phonetics and grammar, alternative assessment seeks to put emphasis on the way the learner ‘uses’ language and literacy (Wrigley, 1992). Chirimbu (2013) asserts that the idea of involving learners in communicative language assessment tasks to gather information about their linguistic level is the biggest motivation of educators to adopt alternative

assessment. In addition, as alternative assessment puts the learner in the center thanks to its flexible and dynamic nature, it has gained tremendous acceptance by learner-centered curriculum developers (Wrigley, 1992). With diverse opportunities, it helps teachers keep their learners motivated, take active roles in the teaching process and thus function as autonomous learners (Rus, 2019).

The increasing use of more complicated technology also contributes to the spread of alternative assessment (Fox, 2017). Animations, hot spots, e-journals, blogs and wikis are among the popular online assessment opportunities which would be not applicable in traditional assessment settings. These alternative assessment tools not only assess students' level of knowledge but also enhance their learning experiences by developing other skills. While traditional assessment is often limited to lower-cognitive skills, alternative assessment goes beyond that by involving team activities and reflection (Robles & Braathen, 2002). To illustrate, Crisp (2011) suggests that teachers could create assessments embedded in virtual worlds such as *Second Life* to improve learners' problem-solving skills. Commonwealth of learning specified the aims of self-assessment and peer-assessment, which are among the most recognized alternative assessment types as

to increase student responsibility and autonomy; achieve a more advanced and deeper understanding of the subject matter, skills and processes; elevate the role and status of students as assessors; encourage a deeper approach to learning; involve students in critical reflection; and develop in students a better understanding of their own subjectivity and judgement. (Westhuizen, 2016, p.15)

Electronic portfolios are one of the best examples of alternative assessment methods which benefit the active participation of learners. E-portfolios are beneficial

in improving assessment turnaround time, student motivation and consistency of quality assessment (Crisp, 2011). They are reflective, collaborative, and creative tools to present knowledge, learning and growth over time. It also gives the teachers a chance to track their students' progress easily and timely. According to Westhuizen (2016), the use of e-portfolios is aligned with the "create" level in Bloom's revised taxonomy as learners are expected to create, construct, design, invent and produce. Robles and Braathen (2002) also highlight the effectiveness of e-portfolios as they make possible the accumulation of homework, assignments and projects, instructor comments. In this respect, online platforms such as Google Sites or Weebly are widely used to encourage students to construct, customize and present their understanding and reflection of the learning objectives.

The effectiveness of alternative assessment has been studied by the researchers for several decades. For instance, Elbow and Belanoff (1986) conducted a study at State University of New York at Stony Brook where the proficiency examination was replaced by portfolio assessment in an English writing course. They assigned students writing tasks and gave them a chance to revise their work upon feedback from their teachers. The results of the study showed that the more feedback students got and revised their work, the higher the passing rate was. It was also found in the study that alternative assessment encouraged peer-feedback and collaboration among students. Another relevant study was conducted by Jia et al. (2006) with an aim to explore the perceptions of English language teachers on alternative assessment of English reading skills. In their qualitative study, the researchers conducted classroom observations, analyzed assessment materials used in classrooms, and carried out interviews with teachers. The findings suggested that all English language teachers perceived the use of alternative assessment in reading to

be more efficient than the use of traditional methods. Another study was carried out by Yildirim and Orsdemir (2013) in Turkish context. The researchers worked with 43 English teachers from 10 different schools in order to explore their perceptions of alternative assessment in language teaching through surveys and interviews. It was found out that most teachers believed alternative assessment methods help students improve their language skills as well as increase student motivation and confidence. One recent study conducted by Farahian et al. (2021) included 69 university students. Through questionnaires and semi-structured interviews, the researchers concluded that portfolios could foster creative thinking and increase the quality of reflection in English language classes. Another study conducted by Fathi and Rahimi (2022) focused on the effect of electronic portfolio tasks on learners' English writing performance. The researchers adopted a sequential explanatory mixed-method approach in their study with two groups of English learners. The results showed that the group who used electronic portfolio tasks performed better than those who used traditional portfolio tasks, and students benefited from electronic portfolio due to its practicality.

Integrated Assessment

Another approach that has gained attention especially with its use in large-scale tests such as TOEFL IBT is integrated assessment. Rather than testing each ability of test takers such as speaking or writing in a second language, integrated assessment includes a combination of skills such as reading-writing tasks, listening-speaking tasks or even reading-listening-speaking tasks. It is usually designed as tasks where the test taker makes use of some source such as a reading text or listening audio that is provided in order to perform another task such as writing a text or giving a speech (Plakans, 2015).

According to Cumming (2013), the advantages of integrated tasks are offering real-life experiences, providing solutions to test method effects related to traditional test items, involving ‘construction-integration’, having learners engage in specific content performances and empowering instruction and self-assessment. Integrated assessment is on the rise also thanks to its potential integration of technology-enhanced assessments. To illustrate, simulation-based assessments are used both in high-stake and classroom assessment in ‘purposeful, interactive and strategic’ methods (Llosa, 2017). All these benefits integrated assessment offers attracted researchers’ attention and thus a lot of research is being conducted on integrated assessment. All but one interview participant of a study which examined test takers’ feedback on integrated assessment stated that this type of assessment is preferable because it supplies the topical source for the test taker and thus lowers the anxiety (Huang & Hung, 2017). In a different mixed-method study, researchers (Sadeghi & Rahmati, 2017) compared two groups of language learners as integrated assessment and non-integrated assessment groups as well as consulting test takers’ and teachers’ notes to have a broader understanding. The quantitative part of the research concluded that the integrated assessment group performed notably better in writing than non-integrated assessment group. In addition, focus groups highlighted that autonomy, independence and meta-cognition were supported through an integrated assessment process. The same study, on the other hand, also shed light on some issues related to integrated assessment such as the complexity of scoring and rubrics, which would inhibit the quality criteria of both teachers and students. In other words, integrated assessment is subject to perils such as lack of precise definition, practicality in scoring and comparability with other levels of language skills (Cumming, 2013). However, workshops on assessment processes along with

some good and bad samples for both stakeholders could be used to overcome these drawbacks (Sadeghi & Rahmati, 2017). And it is crucial that test takers are given guidance and practice in forming correct, accurate and fluent responses to multi-skill integrated assessments (Plakans et al., 2019). Almost a decade ago, Yu (2013) also called for a more comprehensive theoretical framework so that all stakeholders have a better understanding of what, how and why to integrate assessment tasks. This need still maintains its existence as the future designs of assessment continue to refer to a substantial number of tasks that integrated textual and/or aural information provided ahead of time (Huang & Hung, 2017).

Assessment of Language Proficiency

Language proficiency tests are widely used as a criterion for students' eligibility to enroll in and graduate from certain education programs as well as providing diagnostic results for various decision-making processes (Wolf & Faulkner-Bond, 2016). Immigration and globalization are only two of many reasons why the number of English as a second language learners is getting bigger every year (Llosa, 2017). As the population of English language learners is increasing, the responsibilities and expectations of educational settings also increase (Guzman-Orth et al., 2016; Miley & Farmer, 2017). The heterogeneity in language proficiency tests brings about many challenges for quality, valid, reliable and ethical proficiency tests. In their suggested approach to language proficiency assessment, Gholson & Guzman-Orth (2019) emphasize that language proficiency tests should be fair, relevant, aligned and informative. Moreover, accessibility should be ensured especially for first-time and young test takers and learners with disabilities (Guzman-Orth et al., 2016).

In response to different conceptualizations of language proficiency tests, Poza and Valdés (2017, p.432) affirm that “English language proficiency tests (1) have different performance levels and (2) test different skills, which are described and measured differently”. The authors also remind that therefore one educational setting’s classified level might totally differ in another setting. In addition, validity concerns have been a great focus of research in the past decade. For instance, Wolf & Faulkner-Bond (2016) addressed the representations of academic and social English in English language proficiency assessments through a linguistic content analysis of assessment items collected from three states. The authors had the assumption that these proficiency tests must be intended to measure academic and social English as to guarantee test takers’ readiness to be active participants in academic contexts. The study concluded that although differing in level and type, all three states contained test samples that would reflect academic and social English across the four language domains (listening, reading, speaking, and writing). Collier and Huang (2020) analyzed the Texas English Language Proficiency Assessment System (TELPAS), which is a high-stake exam conducted annually for school admissions and other language assessment decisions. Intended to measure social and academic English skills in four skills, TELPAS uses a range of test items such as multiple-choice, narrative, interactive and more types. It uses holistic rubrics to score productive skills, and categorizes learners as Beginner, Intermediate, Advanced, and Advanced High. The authors took the following criteria into account while evaluating TELPAS: “reliability, validity, authenticity, practicality, washback/consequences, and test bias and fairness” (Collier & Huang, 2020, p.224). The findings pronounced the importance of sufficiency of the proficiency tests

delivered a year, inter-rater reliability, and supportive communication between stakeholders.

All in all, one of the primary purposes of language assessment is observing achievement and progress, which will also give an idea to detect areas for improvement (Davidson, 2018). It is not only curriculum developers' job to do this; instead, all stakeholders must be actively involved in the design and application of assessment processes. As Taylor (2013) suggests, the major stakeholders in language assessment are test writers, class teachers, administrators and test developers. All of these stakeholders should have different levels of assessment literacy regarding knowledge of theory, principles and technical skills. In other words, assessment literacy will have an impact on how successful new assessment policies are conducted (Jiang, 2020).

Assessment Literacy

As assessment for learning requires teachers to be involved in the assessment procedures more actively than in assessment of learning, teachers are expected to be aware of the principles of assessment literacy to effectively make use of it. Teachers' assessment literacy is a "joint property" (Coombe, 2020) that is dynamic and interacting among different stakeholders such as students, test developers and administrators. As Stiggins (2014) states, students need assessment literacy to understand the purpose of assessments and the scoring guides, to interpret the difference between good and bad performances, and to perceive how the assessment results could be used to support their learning. Students are also concerned with assessment literacy to diminish the negative effect of assessments such as exam anxiety and to actively take part in assessment related processes with increased motivation (Djoub, 2017).

Stiggins (as cited in Green, 2018) proposed three levels of assessment literacy: functional level, practical level and advanced level. The functional level is needed by the professionals who only use the assessment results for pass-fail decisions or ranking. The practical level appeals to the teachers who use and produce assessment data for their students. The advanced level goes beyond these and is needed by assessment specialists who produce assessment data for other stakeholders as well. The current situation both in Turkish and international contexts shows us that assessment training programs usually refer to the functional level addressed by the public officials to use the data; and the advanced level which is practiced by the measurement specialists generating the data. However, the practical level is usually neglected and results in missing the benefits that assessment literacy could provide teachers with. Green (2018) also highlights the need for teachers' assessment literacy by referring to the responsibility of teachers to take a role in both assessing their learners in the classroom and preparing them for other assessments.

The key components of assessment literacy include various items as presented by different experts in the field. As the pioneer of the term, Stiggins (1999) defines seven competencies for assessment practices. These are listed in the authors' article (as cited in Herrera & Macías, 2015) as

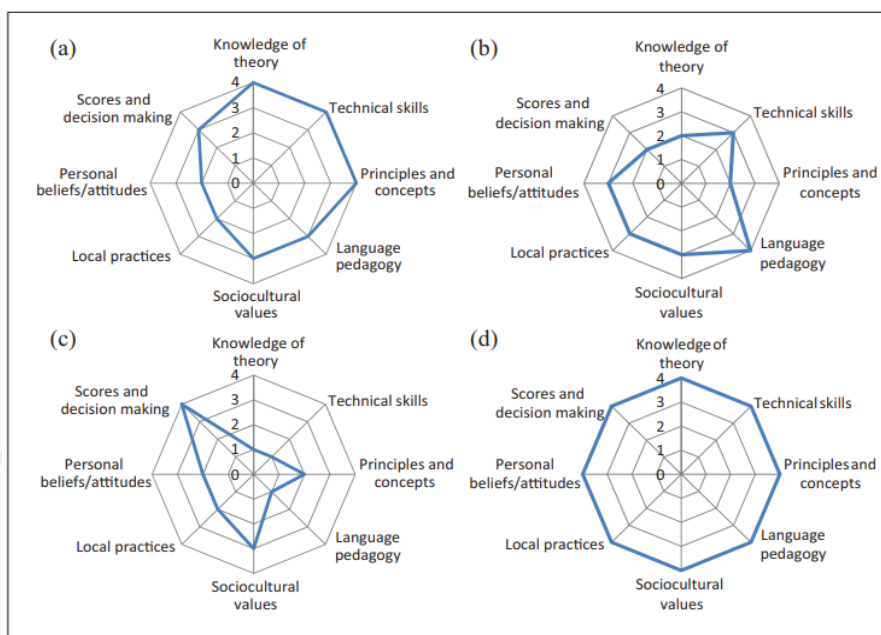
- (a) connecting assessment to clear purposes, (b) clarifying achievement expectations, (c) applying proper assessment methods, (d) developing quality assessment exercises and scoring criteria and sampling appropriately, (e) avoiding bias in assessment, (f) communicating effectively about student achievement, and (g) using assessment as an instructional intervention. (p. 307)

As the concept assessment literacy is not as old as the assessment term itself, many teachers are not assessment literate (Popham, 2018). Despite the awareness-

raising attempts in the language programs over the past few years, assessment literacy is not viewed as an opportunity to benefit teaching practices. Even though there are some training courses specially designed to increase assessment literacy both in pre-service and in-service training of EFL teachers (Öz & Atay, 2017), these courses are usually based on large-scale testing and statistical analyses instead of in-class assessment practices. Since the term ‘literacy’ has gone beyond just ‘the ability to read and write’ (Taylor, 2013), assessment literacy has also needed to adopt a wider notion embracing sociocultural as well as functional principles. In addition, the growing number of social domains requiring a heightened awareness and knowledge of assessment has led to varying degrees of assessment literacy for different stakeholders (Kremmel & Harding, 2020). Taylor (2013) calls for the integration of assessment development activities into professional development programs with relevance to a range of stakeholders. She states that all stakeholders should be provided with the core expertise in language assessment literacy to better inform and shape innovative policy making and practice.

Theoretical Framework of the Study

The theoretical assumption of the study derives from Taylor’s (2013) framework of suggested language assessment literacy profiles for four constituencies, which can be seen in Figure 1 below. This framework was chosen for the current study because the participants of the study also included English language professionals with different profiles, such as teachers, assessment specialists, and academic leaders.

Figure 1*Differential Language Assessment Literacy Profiles*

Note. (a) Profile for test writers, (b) Profile for classroom teachers, (c) Profile for university administrators, (d) Profile for professional language testers. Differential language assessment literacy profiles. From “Communicating the Theory, Practice and Principles of Language Testing to Test Stakeholders: Some Reflections,” by L. Taylor, 2013, *Language Testing*, 30(3), p.410

(<http://dx.doi.org/10.1177/0265532213480338>). Copyright 2013 by Sage.

As illustrated in Figure 1, Taylor (2013) asserted that different stakeholders require different assessment knowledge and skills based on their job titles and responsibilities. Professional language testers are the ones who need literacy in all of the specified dimensions. For *teachers*, the most important dimension is *language pedagogy*, while this dimension is regarded as the least important one for academic leaders. It can also be seen that *university administrators* (academic leaders in the current study) need the most assessment literacy in *scores and decision-making*. As for the *test writers* (assessment specialists in the current study), *knowledge of theory*,

principles and concepts, and technical skills are the most important assessment literacy dimensions as specified in Figure 1.

With an attempt to investigate Taylor's (2013) proposed profiles, and design an assessment literacy survey at the service of stakeholder groups, Kremmel and Harding (2020) conducted a large-scale study with different stakeholders. In a 24-month research and development time, they extended the hypothesized dimensions so that the survey would be feasible and intelligible to different populations. Through an empirical investigation of the language assessment literacy needs of different stakeholders, the researchers came up with a comprehensive survey to demonstrate levels of language assessment literacy across different stakeholder groups. After going under four revisions, the final version of the survey was sent to 2419 participants (language teachers, professional examiners, test developers, testing researchers, policymakers, test score users, test takers, and parents of test takers) all around the world, but 1086 surveys- mostly language teachers (645), followed by language assessment/test-developers (198) and language assessment/testing researchers (138) were used after the cleaning of incomplete and low-confidence responses. The researchers asked the participants "How knowledgeable/ skilled do people in your chosen group/profession need to be about each aspect of language assessment?" and carried out exploratory factor analysis of the hypothesized items for eight dimensions. Following the necessary statistical analyses, the authors validated their survey of 53 items (initially the number was 74 but 21 items were removed after the factor analysis) to measure assessment literacy under ten dimensions, subdividing the technical skills into three dimensions (constructing language assessments, administering and scoring language assessments, evaluating language assessments). The results validated the notion in Taylor's (2013)

framework that different profiles need different set of skills and knowledge based on their job descriptions. Therefore, the tool was thought to be feasible for the current study conducted with different profiles.

Assessment Literacy for Teachers' Development

In order to determine teachers' level of assessment literacy, some tools have been suggested in the literature. An older version of these tools is the 'Classroom Assessment Literacy Inventory' developed by Mertler and Campbell (2005). The inventory has 35 items in total which aim to measure assessment literacy under seven headings "connecting assessments to clear purposes, clarifying achievement expectations, applying proper assessment methods; developing quality assessment exercises and scoring criteria and sampling appropriately, avoiding bias in assessment, communicating effectively about student achievement, using assessment as an instructional intervention" (p. 7). As another tool, Fulcher (2012) developed and applied a survey to address the assessment training needs of language teachers. The survey explored teachers' perceived assessment needs within three categories as practical knowledge, theoretical and procedural knowledge, and socio-historical understanding. Another recent tool was developed by Nikmard and Zenouzagh (2020). Through extensive literature review and interviews with Ph.D. candidates in English language teaching, the authors came up with an assessment literacy questionnaire within a theoretical framework for four main components of assessment: validity, reliability, interpretability of the results and efficiency. Especially when the impact of assessment has been acknowledged across different stakeholders- not only teachers themselves- around the world, scholars and educational organizations initiated attempts to define different levels of assessment literacy for specific stakeholders. For example, the Michigan Assessment

Consortium (MAC) updated and released a set of Assessment Literacy Standards written for five groups of individuals: students and their families, classroom teachers, building administrators, district administrators and policymakers. Intended to contribute to a more assessment-literate population, the standards for each group address three domains: dispositions, knowledge and performance.

It was also found out in a case study that assessment literacy impacted teachers' classroom questioning and practices at personal, institutional, and socio-cultural levels which are shown in Table 1 below (Jiang, 2020).

Table 1

Factors Influencing Teacher Assessment Literacy in Questioning Practice

Levels	Influencing Factors
At the personal level	Teacher expertise in assessment Teacher beliefs
At the institutional level	Student competence and motivation Institutional value
At the socio-cultural level	Confucian influenced culture

Note. Adapted from “Teacher Classroom Questioning Practice and Assessment Literacy: Case Studies of Four English Language Teachers in Chinese Universities,” by Y. Jiang, 2020, *Frontiers in Education*, 5(23), p. 10 (<https://doi.org/10.3389/feduc.2020.00023>). Copyright 2020 by Frontiers in Education.

Another study that links assessment literacy and professional development was conducted by Bijsterbosch et al. (2019). Through questionnaires, interviews and content analysis of summative assessments, the researchers found out that reflection on educational goals, beliefs and feedback contributed to higher levels of assessment literacy. Sheean (2017) states that there is an interconnection among autonomy,

teaching methods and assessment methods. Autonomy and assessment relation is definitely not limited to learner's autonomy when assessment literacy itself requires the active participation of the teachers in the processes. The author argues that although much research has been done on the advantages of taking responsibility in learning processes for high autonomy, the potential benefits of active participation in the assessment procedures are usually neglected.

The neglected benefit of assessment literacy among teachers was researched in a study which showed that 81.39% of teachers surveyed said that most of the assessment practices are carried out by specific specialists at their institutions and they, as teachers, have almost no word in these decisions or practices. Although these teachers have some ideas and assumptions about assessment thanks to some courses or trainings they took, they did not have a chance to improve and update themselves in terms of key concepts and practical applications of quality assessment (Djoub, 2017). Therefore, it becomes almost impossible to increase teacher autonomy in assessment without providing them with the necessary knowledge and insight to be assessment-literate.

Perceptions of Assessment Literacy

Higher quality education is perhaps the ultimate goal and desired outcome of assessment literacy. Designing, applying, and interpreting correct assessment tools will help reach educational and institutional objectives. As Green (2018) suggests, assessment does not only mean certain numbers or grades for statistical analyses, but instead should add to the development of teaching and learning. Therefore, key stakeholders must be able to assess student achievement with the right practices and refer to the results in order to improve their instruction (Shams & Iqbal, 2019).

There have been some studies as to how assessment literacy is perceived by key stakeholders. Tsagari and Volt (2017) carried out a mixed-methods study with 853 questionnaires and 63 interviews with language teachers from selected European countries. The authors aimed to explore teachers' perceptions of assessment literacy in terms of their training backgrounds and needs with a special focus on the reflections of this literacy on their classroom-based practice. The results showed that most of the participants believed that their pre-service or in-service training did not sufficiently prepare them for their assessment-related practices. Moreover, the majority of the participants' ability to design classroom-focused language assessment practices was underdeveloped. It was concluded that this lack of knowledge and skills refrained them from using innovative assessment methods. In the interviews, they stated that although they have a fuzzy understanding of what alternative assessment formats were, they could not feel confident enough to implement them properly in their classes. A final commonly pronounced perception was that most teachers wished for better training on assessment so that they could improve their feedback strategies for a given test or assignment (Vogt & Tsagari, 2014).

A similar study on teachers' perceptions of assessment literacy was carried out in a Turkish context (Öz & Atay, 2017). The researchers conducted semi-structured interviews with teachers at the English language preparatory school of a university in order to discover their assessment literacy perceptions and how much these perceptions were reflected in their performances. The results showed inconsistency between teachers' perceptions and their real performances. For instance, although teachers believed that good assessment must be valid, most of them focused on only speaking skill while assessing their students' English language proficiency.

Another specific research was conducted on Indonesian teachers' perceptions of instruction-based assessment practices and how much these perceptions are reflected in their teaching practices (Zulaiha et al., 2020). The researchers explored English as a foreign language (EFL) teachers' knowledge about assessment principles through a survey and gathered more qualitative data through their curriculum, lesson plans and sample assessment tools. They also conducted semi-structured interviews to determine the extent to which the teachers applied their knowledge in their classroom practices. The results mainly showed that most teachers believed in the importance of using assessment as a tool to promote learning. For instance, they mentioned that teachers should conduct assessments to not only measure success but also get feedback on instruction; and the assessment plan should be well communicated to students by their teachers. However, although teachers demonstrated sufficient knowledge of assessment practice including planning, monitoring and so forth, there were some gaps between what they perceive as and what they do in real classroom settings. To illustrate, although most teachers believed in the importance of feedback to assessment practices, the authors could not find evidence of quality feedback given by the teachers. Overall, it was concluded that school policy, students' attitudes and parents' involvement are some concerns that affect teachers' perceptions of assessment literacy and their in-class assessment practices.

Vogt et al. (2020) investigated perceptions of assessment literacy and assessment literacy levels as well as teacher training needs regarding assessment literacy with a mixed-method comparative study in Germany and Greece. They collected data through questionnaires, and interviews were used for supporting evidence. The study concluded that both Greek and German teachers had limited

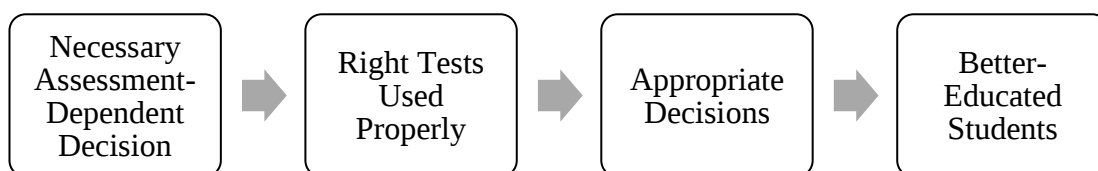
formal training in assessment. Also, both groups had similar perceptions in their assessment conceptualization although they had different forms of assessment practices in their school settings. However, their perceived training needs were quite different from each other. Researchers made it clear that contextual factors such as school characteristics, needs and traditions play a great role in this difference. As a result, professional development activities should be designed in an attempt to increase teachers' assessment literacy levels with relevant contextual factors in mind.

Conclusion

The creator of the assessment literacy term, Stiggins (1991) warns that assessment literacy is often seen as a skill only for test developers as the ones who create the tests and the ones who teach have for a long time been separated. This spreads the perception that assessment and instruction do not intersect (Stiggins, 2014). However, teachers will soon gain the positive outcomes of being literate in assessment, which will improve the quality of education (Popham, 2018). The author elaborates on this by stating that assessment-literate teachers would refrain from using wrong tests, misusing the data, and failing the instructional benefits of assessment literacy. He visualizes the decision impact of assessment-literate educators in Figure 2.

Figure 2

Decision Impact of Assessment-Literate Educators



Note. Adapted from “*Assessment Literacy for Educators in a Hurry (p. 12),*” by W. J. Popham, 2018. Copyright 2018 by ASCD.

Finally, assessment literacy will give the teachers a chance to generate assessment tools which will be in accordance with standardized tests such as TOEFL or IELTS (Herrera & Macías, 2015). Millions of people around the world are taking these international exams to be admitted to higher education programs or certain promotions. As a result, assessment literacy will in a wider context pay off for various key stakeholders involved in assessment processes.

Educators and policymakers should hold the belief that teachers' being assessment literate benefits quality learning and teaching. Accordingly, policymakers should include assessment practices that are purposefully designed to help students raise awareness on the various elements of assessment and thus direct their own learning as well as provide valuable feedback for improvement. When curriculum developers and educators plan programs, courses, materials and so on and so forth with the approach to assessment as learning, this would yield positive results for all stakeholders. To add to these positive results, teachers, as key stakeholders, should develop assessment literacy.

It is clear that being assessment literate and active advocates of assessment as learning cannot be fully achieved without appropriate training. This link between assessment literacy and teacher development was also mentioned by Herrera and Macías (2015). The authors asserted that various professional development activities such as workshops and study groups may be helpful to present relevant and meaningful knowledge to educators, which will contribute to their professional growth. That is to say, assessment should be better integrated into teacher training programs with the specific content for assessment literacy (Djoub, 2017). This will help close the gap between the practices and the principles of assessment.

CHAPTER 3: METHOD

Introduction

In this chapter, the research design is explained with the methods described and justified. The context where the research was conducted is briefly portrayed and the distinctive characteristics of the case setting are outlined. There is also some descriptive information about the participants of the study. Methods of data collection and analysis along with instrumentation are also reported in this chapter.

Research Design

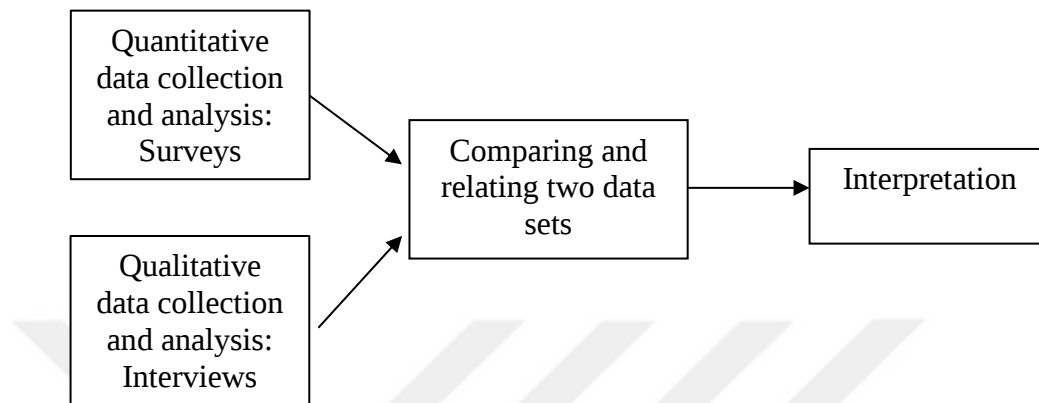
This study is conceptualized as a case study employing the convergence model of mixed-method design (Gay et al., 2009). Mixed-method research is an approach that requires the mixing of quantitative and qualitative data so that the results would yield a more powerful and stronger interpretation of a research problem (Creswell, 2014), the interpretation of perceptions of assessment literacy in this case. This research design was decided to be appropriate as mixed-method research reinforces the reliability of the research by comparing different angles from quantitative and qualitative data as well as providing a more sophisticated approach to the research questions. The nature of the research design necessitates extensive data collection, which is also time-intensive; however, the accessibility of both forms of data made this research method applicable and favorable for this study (Creswell, 2014).

In the convergence model of mixed-method design, the researcher collects data from different sources to confirm quantitative results with qualitative results to end up with strong conclusions about the same phenomena (Creswell & Clark,

2017). In this study, quantitative data from the surveys and qualitative data from the interviews were collected in the same time frame, but separately (Figure 3).

Figure 3

Convergent Mixed Method Procedures



Note. Adapted from “*Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (p. 220),” by J. Creswell & J. D. Creswell, 2014. Copyright 2014 by Sage Publications.

The qualitative data were used to confirm quantitative results. Detailed views of the participants obtained as qualitative data via interviews were used to support the findings of the quantitatively collected and analyzed data coming from the surveys.

Both forms of data focused on the same or parallel concepts, as in this case assessment literacy knowledge and skills specified in Taylor’s (2013) framework. As a typical characteristic of a mixed-method design (Creswell & Clark, 2017), the sample size was smaller for the collection of qualitative data ($N = 21$) than for the quantitative data ($N = 33$). However, this should not necessarily be considered as a drawback as both research designs differ in their intention. After that, two data sets were merged for an overall interpretation to answer the research questions. To better understand the phenomena in the study, data sets were compared and the

interpretation of both was reported; thus, the study is a convergence model of triangulation design (Creswell & Clark, 2017). Triangulation refers to the act of combining several research methods to explore and validate the results of the same phenomena (Ndanu & Syombua, 2015). Methodological triangulation is suggested by the relevant literature as it establishes validity in mixed method research designs (Ma & Norwich, 2007).

Case studies can be described as particularistic, descriptive and heuristic (Gay et al., 2009). This study was carried out in a bounded system, at the English Language Preparatory school of a foundation university in Ankara, Turkey. Yin (2003) defines a case study as “an empirical inquiry that investigates a contemporary phenomenon within its real-life context” (p.13). This study also explored the English language teachers’ perceptions of assessment literacy within a real context; the institution where they work. Therefore, purposive sampling was adopted for the better association of the sample with the research questions (Campbell et al., 2020) and to empower the trustworthiness of the results. The inclusion of real-life context and contextual conditions are important in the study as they would affect teachers’ perceptions to a great extent.

Because both assessment literacy and perception are phenomena that comprise deep notions for one to comprehend, thick descriptions were required to answer the research questions. In this study, there have been thick descriptions as a result of surveys and interviews, both of which aimed to bring insights into the phenomenon of assessment literacy. As case studies are designed to explore such phenomena with long narratives, which are usually difficult for the reader, a set of open-ended questions were used in the study as suggested by Yin (1981).

Triangulation, employed to ensure the credibility of the data, were conducted using

various instruments to ascertain the perspectives of multiple stakeholders, namely teachers, assessment specialists and academic leaders. With this strategy, potential risks and biases were aimed to be reduced (Maxwell, 2008). The use of multiple resources of data ensured cross-validity of the results through the triangulation process (Kaplan & Maxwell, 2005).

Context

The case for the current study is the English language preparatory school of a foundation university with 30% English-medium and 70% Turkish-medium of instruction. The university provides instruction at undergraduate, graduate, and PhD levels. The university accepts around 1000 students for its various discipline of studies in total every year. The English language preparatory school is a part of the Foreign Languages Department and is managed by the department head. The department also offers second foreign languages such as Italian language and departmental English courses, but they are out of the scope of this study. The English language preparatory school of the university where this study was conducted is an authorized center for internationally recognized English language proficiency tests such as TOEFL, TOEIC, and GMAT.

Assessment Regulations

All students who are admitted to the university take the preliminary English language exam every year. Those who fall below the cut-off score enroll in the English language preparatory school (around 600 students). These students are taught by around 40 full-time and 25 part-time English language teachers.

The preliminary exam is an in-house test of the English language prepared by the Measurement and Evaluation Unit of the preparatory school. The exam has four parts: listening, grammar, vocabulary and reading. Students who score between

0-49 start studying in the prep school at the lowest level: A Foundation (AF). The ones that score between 50-64 start at A level. These two groups of students cannot take the TOEFL ITP yet to be exempted from studying in the prep school. However, the students who score 65 and above in the preliminary test are eligible to take the official TOEFL ITP, which is the proficiency exam of the school. Those who score 500 out of 677 from the TOEFL ITP can directly continue their studies in their departments. On the other hand, if they score between 400-449, they start the prep school at B level; and if they score 450-499, they are placed in the highest level: C level.

In conclusion, there are four main levels, AF, A, B, and C levels. There are also repeat versions as A Repeat (failing students from A and AF levels), B Repeat and C Repeat. By the end of the semester, the grade points average (GPA) based on the components in the assessment schedule must be 70 for AF and A levels, 65 for B levels in order for the students to pass onto the next level. If they cannot, they take the repeat version of the same level. A regular student must complete C level with a GPA of at least 65 points to be eligible to take the TOEFL ITP exam. However, the institution gives this early right also to B level students with an GPA of 80 points and above.

Assessment Schedule

Since the foundation of the preparatory school in 2005 until the pandemic in 2020, students had to take 11 multiple choice exams in the form of quizzes and midterms, four writing exams, and one speaking exam during their study of each level in a semester. There were also some online components such as online homework of course books in the assessment schedule. However, as of the 13th of March 2020, after the breakout of the Covid19 pandemic, the school had to change

its assessment system. With the introduction of more formative types of assessment tools, the institution has been in an attempt to and is still extending the use of assessment for learning.

Unlike the system before the pandemic, at the time of the study, there were fewer multiple-choice type examinations. Instead, formative assessment tools in which students worked on several tasks almost every week and revised their work upon feedback from teachers were aimed to be conducted. In addition, all students completed online homework on three different platforms to prepare for the coming week's topics as well as revision activities for the previous week's topics. There were two midterm exams, three quizzes and a final exam in the form of multiple-choice questions. These exams constituted 71% of the final grade. Ten percent of the cumulative grade came from two portfolio tasks; 9% from a speaking exam, and 10% from online assignments of course materials.

The Unique Setting

This preparatory school, within the department of foreign languages, was chosen for this case study for a reason. As highlighted by Yin (2003), case study research deals with technically distinctive settings where the relevant data is more than just numbers. The current study took place at this preparatory school that has a distinct setting. There are two reasons for this distinction. First, the school is different from its counterparts because it follows a unique curriculum. The academic calendar is composed of trimesters each of which lasts 13 weeks. It also requires the students to pass an internationally accepted and preferred English test as its proficiency exam. Secondly, there is a distinctive emphasis on assessment issues in the institution. This emphasis is often pronounced by the head of the department who has an MSc and a Ph.D. degree in assessment and evaluation. All the assessment

practices are designed and conducted by a separate unit, the Measurement and Evaluation Unit which has five members managing different levels. The members are responsible for planning the assessment schedule, designing appropriate assessment tools, administering assessments and coordinating the marking and feedback processes. All the relevant procedures that are followed by teachers are supported by checklists, rubrics, regular meetings, and student information systems. Confidentiality is a noticeable principle when it comes to certain assessment procedures. To illustrate, under no circumstances should the students be informed that the exams are prepared by the teachers working in a specific unit.

The attention given to assessment also shows itself in a recent training program on assessment offered to the teachers as part of professional development activities. The training was voluntary for the teachers except for the members of the Curriculum and Material Development Unit who were obliged to attend the sessions as told by the head of the department. Offered as a Peer-learning project of professional development practices, the training was designed by the members of the Measurement and Evaluation unit and the head of the department before the pandemic. The purpose of the training was to increase teachers' knowledge and skills in assessment. There were eight input sessions in total, each of which took around 50 minutes: (1) Fundamentals of Testing, (2) Designing Assessment Tools, (3) Assessment of Use of Language, (4) Assessment of Reading, (5) Assessment of Listening, (6) Assessment of Speaking and Writing, (7) Alternative assessment, and (8) Item Analysis. Two of the sessions were presented by the head of the department himself and the members of the Measurement and Evaluation Unit presented the rest. The first five sessions had been completed face to face before the pandemic. Then, because of the sudden shift to online teaching, the training had to be halted for a few

months. Later, two sessions were conducted online. And the last session was conducted face-to-face in the 2021-2022 academic calendar. There were 12 teachers who attended all of the sessions and the number was higher in a few of the sessions. The major points covered in the course were principles of assessment, item writing guidelines, and using statistics to evaluate assessments.

It can be said that the pandemic brought along more formative assessment components to the schedule such as the portfolio tasks. However, the midterms, the final exam and the proficiency exam were still in the form of multiple-choice type of questions. In addition, although assessment training was provided for volunteer teachers, these teachers did not take active roles in the design and evaluation of the assessments. As there is a special unit (Measurement and Evaluation Unit) for all types of assessment processes, other teachers have very little to say in the design, implementation, and evaluation of them. Their role in assessment processes is often limited to proctoring the exams and marking students' work with the provided rubrics. They may also provide feedback to the unit in the weekly coordination meetings or through surveys filled in at the end of each semester. The lack of chance to be actively involved in assessment processes may strain assessment literacy and result in a particular direction in their perceptions of assessment literacy and its relationship to the teaching quality or assessment for learning.

Participants

Data were collected from English language teachers, assessment specialists and academic leaders at the institution who volunteered to participate in the study. Thirty-three English language professionals completed the survey in total. Twenty-one of them volunteered to be interviewed as well. Table 2 shows the demographic information of the participants who completed the survey.

Table 2*Demographic Information of English Language Professionals (Survey)*

Variable	Category	<i>n</i>
Gender	Female	28
	Male	6
	Total	34
Age	21-25	3
	26-30	6
	31-35	15
	36-40	4
	40+	6
	Total	34
Title	Teacher	23
	Assessment specialist	4
	Academic leader	7

As can be seen in Table 2, the majority of the participants are females aged between 31-35. In total, 23 language teachers, four assessment specialists and seven academic leaders completed the survey. *Language teachers* are responsible for teaching 20 hours a week. They also mark student papers such as writing sections of the exams and portfolio tasks such as presentation slides.

Assessment specialists are the teachers who work in the Measurement and Evaluation unit as well as teaching ten hours a week. They are responsible for designing, creating, applying, and evaluating all the assessment tools during each semester. During the time of the study, the unit consisted of five people who are experts at English language teaching with either an undergraduate degree in English language teaching or a relevant pedagogical formation. Except for one member with an undergraduate degree, all four assessment specialists had further degrees in different departments. One of the members was pursuing his Ph.D. in Measurement

and Evaluation at the time of the study. Another one held an MA in English language teaching. A third member had an MA in American Literature and Culture. The last member, not a participant of this study, held an MA in Human Resources Development in Education and was pursuing a second MA in Curriculum and Instruction. All of the assessment specialists were trained on the job by the head of the department who has a Ph.D. in Measurement and Evaluation, and by more experienced assessment specialists before they started working in the unit.

Academic leaders are the teachers who work in the administration and as unit coordinators. They are the members of the academic board of the English language preparatory school, meet on a regular basis to take administrative decisions, collate feedback from the units, and make the necessary adjustments related to various issues such as the curriculum and assessment design, extracurricular activities, school-wide changes and so forth. The administration consists of one head, one assistant head, and one coordinator of administrative affairs. During the time of the study, the head of the department held an MA and a Ph.D. in Measurement and Evaluation. The assistant head held an MSc and a Ph.D. in Educational Management and Planning. The coordinator of administrative affairs held an MA in Educational Management and an MA in Curriculum and Instruction. The other four academic leaders were the coordinators of the Curriculum Development Unit, Professional Development Unit, Educational Technologies Unit and Extra-curricular Activities Unit.

The English language professionals who participated in the study had different profiles in terms of years of experience at the institution and their total experience in teaching English. Their levels of education also showed variation to a

certain extent. Table 3 demonstrates the educational and occupational background of the participants who completed the survey.

Table 3

Background Information of English Language Professionals (Survey)

Variable	Category	<i>n</i>
Experience at the institution	0-5	11
	6-10	15
	11-15	4
	16+	4
Experience in teaching English	0-5	6
	6-10	14
	11-15	9
	16-21	1
	21+	4
Level of education	Bachelor's degree	3
	MA/MS student	8
	MA/MS degree	15
	PhD student	5
	PhD degree	3

As indicated in Table 3, 44,2 % of the English language professionals who responded to the survey have been working at the case institution for six to 10 years. Similarly, 41,2% of the participants have had a total of six to 10 years of experience in teaching English. In addition, except for the three participants who have bachelor's degrees, all of them have been involved in further education degrees as shown in Table 3.

Table 4 below displays the demographic information of the participants who took part in the interviewees.

Table 4*Demographic Data of English Language Professionals (Interview)*

Variable	Category	<i>n</i>
Gender	Female	15
	Male	6
Age	31-35	11
	36-40	4
	40+	6
Title	Teacher	10
	Assessment specialist	4
	Academic leader	7
Level of education	Bachelor's degree	3
	Master's degree	13
	PhD student	3
	PhD degree	2

As can be seen in Table 4, the majority of the interviewees were females aged between 31-35. The participants were 10 language teachers, four assessment specialists and seven academic leaders: 21 English language professionals in total. Except for three participants, all of them have been involved in further education degrees.

Instrumentation

The following instruments were used to collect data for the study:

- An online Likert-type assessment literacy survey with three versions; one for language teachers (Appendix A), another for assessment specialists (Appendix B) and the last for academic leaders, (Appendix C)
- A set of interview questions for English language professionals (Appendix D).

Surveys

The survey was originally developed by Benjamin Kremmel and Luke Harding in 2017 and the authors' model was published in 2020. While they developed this survey, as a starting point, they referred to Taylor's (2013) hypothesized dimensions of language assessment literacy; specifically: 1) knowledge of theory, 2) technical skills, 3) principles and concepts, 4) language pedagogy, 5) sociocultural values, 6) local practices, 7) personal beliefs/attitudes, and 8) scores and decision making (Kremmel & Harding, 2020).

To collect some demographic information including age, gender and level of education and experience, the researcher added a section which forms the first part of the survey. The second part, adapted from the original, includes 53 items to explore the perceived assessment literacy level of English language teachers. The difference between the three versions of the survey lies in the question directed to the participants. While in the first version, the language teachers are asked to indicate their own assessment literacy levels, the assessment specialists and the academic leaders are requested to indicate how they perceive English language teachers' assessment literacy levels.

Table 5 shows the number of items in the survey allocated for each dimension of assessment literacy.

Table 5

Number of Survey Items per Each Assessment Literacy Dimension

Dimension	Number of Items
Language pedagogy	11
TS (A) Constructing language assessments	8

Table 5 (cont'd)*Number of Survey Items per Each Assessment Literacy Dimension*

Dimension	Number of Items
Principles and concepts	7
TS (C) Evaluating language assessments	7
Knowledge of theory	4
Local practices	4
Personal beliefs and attitudes	4
TS (B) Administering and scoring language assessments	4
Impact of social values	3
Scores and decision making	1
Total	53

Note. TS= Technical Skills

The first 33 items refer to the “knowledge” of the participants, whereas the rest refers to the “skills” of the participants. The survey is a Likert-type survey which asks the participants to rate the items by assigning a score between one to five; one meaning “not knowledgeable/ skilled at all” and five meaning “extremely knowledgeable/ skilled”.

For this study, the survey was adapted to a very small extent before the piloting process with four English language professionals with the aim of identifying any problems before implementing it to the actual participants. Expert opinion was sought from both the advisor and an assessment specialist for the items. The feedback was evaluated by the advisor and confirmed to revise and finalize the instruments. Considering local needs, items 8, 9, 27 and 28 were adapted to make the

meaning clearer to the participants. For items 8 and 9, to better explain and exemplify the “local systems of accreditation”, “YÖKAK” and “DEDAK” were added to the items to exemplify the relevant local accreditation bodies. For items 27 and 28, instead of saying “local area/context”, “Turkish preparatory school” expression was used to make the participants better understand the boundaries of the study.

To measure the internal consistency of the survey applied in this study, Cronbach’s alpha value was calculated for the survey as a whole, for knowledge and skills categories, and for each dimension in the survey. The Cronbach’s Alpha value for the whole survey was .95, which demonstrates a strong level of reliability for the research (Fraenkel et al., 2012). In addition, the Cronbach’s alpha levels for two main categories (knowledge and skills) were calculated. The Cronbach’s alpha level for the knowledge related dimensions was .93, and the level was .93 for the skills related dimensions. Furthermore, the Cronbach’s alpha values for each dimension in the survey were calculated, and the levels are presented in Table 6 below.

Table 6

Cronbach’s Alpha Values for Each Assessment Literacy Dimension in the Survey

Dimension	Cronbach’s Alpha
Knowledge of theory	.78
Language pedagogy	.87
Personal beliefs and attitudes	.82
Local practices	.80
Impact of social values	.61
Scores and decision making	-

Table 6 (cont'd)*Cronbach's Alpha Values for Each Assessment Literacy Dimension in the Survey*

Dimension	Cronbach's Alpha
Principles and concepts	.83
Constructing language assessments	.90
Administering and scoring language assessments	.84
Evaluating language assessments	.89

Note. The Cronbach's Alpha value could not be calculated for the scores and decision making dimension as the dimension consists of only one item.

As can be seen in Table 6, the Cronbach's alpha values for all dimensions except for impact and social values are all above .75, which suggests a fairly high level of internal consistency of the data collection tool (Gay et. al, 2009). The value for impact and social values ($\alpha = .61$) is considered moderate in the sampled studies (Taber, 2018).

Interviews

To support the quantitative data obtained via the surveys for the first research question, and to answer the second research question, qualitative data were collected through open-ended interviews. The semi-structured interview questions were formed in an attempt to gather data about participants' perceived assessment literacy levels and its relationship with teaching quality. Within the scope of Taylor's (2013) framework, the questions were designed to address the participants' different profiles such as teachers, assessment specialists and academic leaders. There were four main questions with some follow-up questions to be used when necessary. They focused on English language professionals' perceptions of assessment literacy levels and its relationship with teaching quality. As the interview results constitute the qualitative

part of the research, open-ended questions aimed to guide the participants to share and reflect as much as possible. The researcher avoided leading questions and included divergent questions to facilitate the collection of interview data (Gay et al., 2009). The interview questions were piloted with four teachers at the institution to ensure that the questions make sense (Fraenkel et al., 2012). Pilot study results showed no need for change in the form and content of the questions. Each interview lasted around 12 minutes and was conducted in English.

Method of Data Collection

Firstly, the required approval from Bilkent University Ethics Committee was obtained as part of the requirements of the MA in Curriculum and Instruction program. Upon getting the approval, a research proposal form was sent to the English language preparatory school of the university where the study was conducted. Required permissions and consent forms were received from the institution and the participants.

Three versions of the online survey were sent to the participants as a link through an e-mail. The first version of the survey was sent to the 53 English language teachers. 23 teachers completed the survey and 10 of them stated in the survey that they would like to be interviewed. The second version was sent to four assessment specialists and the third version to seven academic leaders. They all filled in the survey and volunteered to be interviewed.

Ten English language teachers, four assessment specialists and seven academic leaders were interviewed face-to-face. The interviewees were recorded upon their consent and later transcribed through a software program called *Descript*.

Methods of Data Analysis

The quantitative data from the survey was analyzed through a software program; Statistical Package for Social Sciences (SPSS 24.0). Firstly, three data sets from the three versions of the surveys were merged into the same file. Each group was assigned a group variable (1 for teachers, 2 for assessment specialists and 3 for academic leaders) to conduct analyses at one time, and to present the results in the SPSS outputs in groups. Distributions were run to report participants' level of assessment literacy for each dimension. Means for measures of central tendency and standard deviations for measures of variability were calculated by running descriptive statistics for the survey.

The qualitative data collected in the form of audio recordings of the interviews were first transcribed through a transcription software named *Descript*. The unclear or misspelled words were corrected, and explanations of the abbreviations needed for the nature of the study were added to the transcribed documents by the researcher. First, theoretical categories (knowledge and skills) that were anticipated prior to interviews were used as two main categories (Maxwell, 2008). Then, sub-categories were created based on the dimensions derived from Taylor's framework (2013). There were ten sub-categories in total. The preliminary list of concepts for each sub-category was derived from the survey items. The preliminary list was reviewed and revised by the researcher and a research expert while the researcher was analyzing the transcriptions. A couple of sub-categories and concepts were added during the analysis process as some comments of the participants did not fit under the existing ones such as in the case descriptive comments they made. The added sub-category and concepts were specified with an asterisk in the final indication of coding categories (Appendix E).

The data was organized in excel sheet tabs for each group of participants. Relevant comments were copied to the related column in the excel sheet (Yıldırım & Şimşek, 2016). During the analysis process, it enabled the research to sort the data out based on the categories, sub-categories and concepts to explore the assessment literacy perceptions of English language professionals. Figure 4 illustrates a sample analysis of qualitative data.

Figure 4

Sample Analysis of Interview Data

	A	B	C	D	E	F	G	H	I
1	Main Categories (2)	Knowledge (K)							
2	Sub-Categories (10)	Knowledge of Theory (K- KT)		Language Pedagogy (K-LP)					
3	Concepts (35)	Structure of English K-KT-SE	Learning English K-KT-LE	Student Motivation K-LP-SM	Washback Effect Experiences K-LP-WE	Feedback K-LP-F	In-class Assessment K-LP-IA	Student Achievement K-LP-StA	Comments
4									
5	T5			*					Assessment plays such a big role and there is a busy schedule in addition, and there are weeks when we only give the assessment parts , this happens often.
6	T2			*					I feel more effective as a teacher if I know more of the assessment. Expectations and the assessment requirements. If I know what the assessment tool is, then I feel like I have more clarity and more alignment with the assessment tool and with the curriculum. I think it shapes my priorities for the class.
7	T3			*				*	I like to operate on the principle of backward design with correct lesson planning. Figure it out wherever the students are, that's first. And then I know that they've reached there, which means I figured out that assessment. Then third, I based my instruction on that to help my students become successful.

This means
Teacher 3

An asterisk (*) means the corresponding
comment is related to the sub-
categories specified

As can be seen in the sample analysis in Figure 4, the transcribed interview data were arranged on an excel sheet based on the sub-category and concept each comment was related to. On the first column from the left, each interviewee was given a different number; for instance, Teacher 1 was labeled as *T1*. The comments made by each participant were read carefully from the transcriptions and copied on the corresponding row titled *Comments* on the right side of the excel sheet, and the

related sub-category or sub-categories as well as the related concepts were marked by an asterisk. While analyzing the data based on Taylor's (2013) framework, the comments were sorted by the asterisks in each sub-category and concept. By doing this, the researcher was able to see how many comments were made for each sub-category and each concept and report the results accordingly.



CHAPTER 4: RESULTS

Introduction

The purpose of this study was to explore the assessment literacy perceptions of English language professionals at an English language preparatory school of a foundation university in Ankara. As such, the research questions were as follows:

1. What is the assessment literacy level of English language teachers as perceived by
 - a. themselves,
 - b. assessment specialists, and
 - c. academic leaders
 at the language preparatory school of a foundation university in Ankara?
2. How do
 - a. English language teachers,
 - b. assessment specialists, and
 - c. academic leaders

perceive the relationship between assessment literacy and teaching quality?

Perceptions of tertiary level English language professionals were explored by two means. Firstly, English language professionals (teachers, assessment specialists and academic leaders) completed a survey to measure the perceived assessment literacy levels of English language teachers. Secondly, volunteered participants from all groups were interviewed to elaborate on their perceptions of assessment literacy and its relationship to teaching quality.

Table 7 shows the data sources for each research question to explore English language professionals' perceptions of assessment literacy.

Table 7

Matching of Research Questions with the Data Sources

Research Question	As perceived by	Data Source
1. What is the assessment literacy level of English language teachers?	a. themselves	Survey & Interview
	b. assessment specialists	Survey
	c. academic leaders	Survey
2. How do English language professionals perceive the relationship between assessment literacy and teaching quality?	a. teachers	Interview
	b. assessment specialists	Interview
	c. academic leaders	Interview

In this chapter, both quantitative and qualitative data results obtained from the participants will be presented. Descriptive statistics including the means and the standard deviations will be shown for the whole survey and for two main theoretical categories: *knowledge* and *skills*. The same values will also be presented for each of the ten assessment literacy dimensions under these categories.

The qualitative data results will be presented and compared to the quantitative findings overall and based on assessment literacy dimensions. The results are organized as headings for each research question below.

Perceived Assessment Literacy Level of English Language Teachers

The first research question was answered through quantitative data from the survey and qualitative data from the interviews.

Survey results

As for descriptive statistics, the mean value for measures of central tendency and the standard deviation for measures of variability are demonstrated for each version of the survey in Table 8. For the first 33 items in the survey, participants rated the level of assessment literacy knowledge of the language teachers in their context choosing one to five meaning (1) not knowledgeable at all, (2) slightly knowledgeable, (3) moderately knowledgeable, (4) very knowledgeable, and (5) extremely knowledgeable. The last 20 items asked the participants to choose the level of assessment literacy skills of the language teachers again out of a scale from one to five meaning (1) not skilled at all, (2) slightly skilled, (3) moderately skilled, (4) very skilled, and (5) extremely skilled.

Table 8 below shows the means and standard deviations for the assessment literacy survey completed by three groups.

Table 8

Assessment Literacy Levels of English Language Teachers

Variable	Perceived by themselves (n = 23)		Perceived by assessment specialists (n=4)		Perceived by academic leaders (n=7)	
	M_1	SD_1	M_2	SD_2	M_3	SD_3
Assessment literacy level	3.13	0.53	3.11	0.36	2.87	0.65

As can be seen from Table 8, the means of all stakeholders' perceived assessment literacy level of English language teachers are around average ($M_1 = 3.13$, $M_2 = 3.11$, $M_3 = 2.87$). This suggests that English language teachers are *moderately knowledgeable/skilled* in assessment literacy as perceived by all

stakeholders. In addition, participants' responses had low deviation from the mean value since the standard deviations of the three groups are below 1. This shows that participants' scores are similar to each other.

The highest mean of assessment literacy level comes from the English language teachers themselves ($M_1 = 3.13$). Similarly, the mean of English teachers' level of assessment literacy as perceived by assessment specialists is 3.11. However, academic leaders perceive the English language teachers' assessment literacy level a little lower, with a mean value of 2.87. This shows that English language teachers are perceived as more knowledgeable/skilled in assessment by themselves than by assessment specialists and academic leaders.

The first 33 questions of the survey consisted of items under the "knowledge" category. The last 20 questions measured the level of "skills" in assessment literacy. Table 9 shows the means and the standard deviations of the items under these two main categories.

Table 9

Means and Standard Deviations for Knowledge and Skills Categories

Variable	Perceived by themselves (n = 23)		Perceived by assessment specialists (n=4)		Perceived by academic leaders (n=7)	
	M_1	SD_1	M_2	SD_2	M_3	SD_3
Knowledge	3.40	0.46	3.42	0.44	3.18	0.65
Skills	2.68	0.69	2.62	0.37	2.40	0.75

The survey results show that all stakeholders believed that English language professionals are *more knowledgeable than skilled* in assessment literacy. It is clear in Table 9 that the skills category perceived by all three stakeholders has lower

means than the knowledge category. Besides, standard deviations of all three groups are below 1; therefore, it can be said that data are clustered around the mean and participants' responses have little variance.

As Table 9 indicates, the assessment literacy *knowledge* level of English language professionals is *above average* as perceived by all stakeholders. The means are 3.42, 3.40 and 3.16, as perceived by the teachers themselves, by assessment specialists and by academic leaders respectively. English language teachers and assessment specialists have very similar means ($M_1 = 3.42$, $M_2 = 3.40$) of perceived level of knowledge in assessment literacy, although academic leaders have a little lower mean ($M_3 = 3.16$). However, as a result of the surveys, it can be said that all stakeholders perceived English language teachers as *moderately knowledgeable* in assessment literacy.

The level of assessment literacy *skills* of English language teachers as perceived by all three groups is *below average*. The means are 2.68, 2.62, and 2.40, as perceived by the teachers themselves, by assessment specialists and by academic leaders respectively. All standard deviations are below 1, so it can be concluded that the variance among participants' responses was low, and the responses were similar to each other. The slight difference across the perceptions of the teachers, the assessment specialists and the academic leaders shows itself in the perceived level of skills category as well. While the English language teachers and assessment specialists have very similar means ($M_1 = 2.68$, $M_2 = 2.62$), academic leaders perceive English language teachers' level of skills as lower with a mean value of 2.40. This shows that academic leaders perceive English language teachers as a little less skilled than do assessment specialists and teachers themselves. Nevertheless, the survey

results indicate that the English language professionals are perceived as *slightly skilled* in assessment literacy by all three groups.

Table 10 below can be referred to for descriptive statistics of the ten dimensions in the survey.

Table 10

Descriptive Statistics for the Dimensions in the Assessment Literacy Survey

Dimension	Perceived by themselves (n = 23)		Perceived by assessment specialists (n=4)		Perceived by academic leaders (n=7)	
	M_1	SD_1	M_2	SD_2	M_3	SD_3
Knowledge of theory	4.14	0.47	4.44	0.55	4.04	0.64
Language pedagogy	3.63	0.58	3.30	0.70	3.48	0.57
Principles and concepts	3.48	0.83	3.75	0.69	3.00	0.96
Impact of social values	3.01	0.53	3.25	0.17	2.48	0.79
Local practices	2.72	0.65	2.96	0.47	2.61	1.04
Personal beliefs and attitudes	3.33	0.71	3.69	0.52	3.11	0.72
Scores and decision making	3.91	0.67	2.50	1.29	3.57	0.79
TS (A) Constructing language assessments	2.59	0.70	2.56	0.84	2.23	0.72
TS (B) Administering and scoring language assessments	3.10	0.82	3.50	0.48	3.31	1.08
TS (C) Evaluating language assessments	2.48	0.79	2.07	0.58	1.96	0.85

Note. TS = Technical skills

As evidenced by the survey results, the assessment literacy dimension with the highest mean value as perceived by all three groups is the *knowledge of theory* dimension. All stakeholders thought that English language professionals' assessment literacy level of knowledge of theory in assessment literacy is *above average*. The means are 4.14, 4.44, and 4.04 as perceived by English language teachers themselves, assessment specialists and academic leaders in order. The standard deviations for all three groups are also lower than 1, which suggest that the participants' responses clustered around the mean.

The dimension with the second highest mean value perceived by English language teachers themselves is *scores and decision making* ($M_1 = 3.91$). Academic leaders also agreed that English language teachers are *quite knowledgeable* in scores and decision making ($M_3 = 3.57$). On the other hand, the mean value for this dimension as perceived by assessment specialists is much lower ($M_2 = 2.30$). However, a standard deviation above 1 ($SD_2 = 1.29$) may mean that there was no full agreement among assessment specialists.

Language pedagogy is the dimension with the third highest mean value ($M_1 = 3.63$) perceived by English language teachers. This suggests that English language teachers perceive themselves *quite knowledgeable* in language pedagogy. Academic leaders and assessment specialists also agreed with this finding as their mean values for this dimension are also considerably high 3.48 and 3.30 respectively. The standard deviations for all groups are below 1; as a result, it can be said that participants' responses for the language pedagogy dimension were similar to each other.

Furthermore, it can be referred from Table 10 that the dimension with the lowest mean as perceived by all stakeholders is *evaluating language assessments*.

The means are 2.48, 2.01 and 1.96 as perceived by English language teachers themselves ($SD_1 = 0.79$), assessment specialists ($SD_2 = 0.58$) and academic leaders ($SD_3 = 0.85$) respectively. Therefore, results suggest that all groups agree with each other on that the English language teachers' skills in evaluating language assessments are *below average*.

Constructing language assessments is the second dimension with the lowest mean value as perceived by English language teachers, assessment specialists and academic leaders with mean values of 2.59, 2.56, 2.23 respectively. There was agreement across the perceptions of the stakeholders as the skills of teachers in constructing language assessments are *below average*.

To interpret the survey results more in detail, descriptive statistics for the items in two main categories of the survey (*knowledge and skills*) are presented in Table 11 and Table 12.

Table 11

Means and Standard Deviations for Assessment Literacy Levels: Knowledge Items

Items in the survey	Perceived by themselves (n = 23)		Perceived by assessment specialists (n=4)		Perceived by academic leaders (n=7)	
	M_1	SD_1	M_2	SD_2	M_3	SD_3
Item 1	3.39	0.66	3.00	0.82	3.00	0.82
Item 2	3.57	0.90	3.75	1.26	3.14	0.90
Item 3	3.48	0.73	2.75	0.96	3.29	1.11
Item 4	3.09	1.08	2.75	0.96	3.71	0.49
Item 5	3.78	0.74	2.25	0.96	2.86	0.90
Item 6	3.43	0.90	3.00	1.63	3.00	0.82
Item 7	3.91	0.67	2.50	1.29	3.57	0.79
Item 8	2.09	0.95	2.00	0.82	2.14	1.21
Item 9	2.45	1.06	3.00	1.41	2.14	1.07

Table 11 (cont'd)*Means and Standard Deviations for Assessment Literacy Levels: Knowledge Items*

Items in the survey	Perceived by themselves (n = 23)		Perceived by assessment specialists (n=4)		Perceived by academic leaders (n=7)	
	M_1	SD	M_2	SD	M_3	SD
Item 10	3.57	0.95	4.00	0.00	3.29	1.50
Item 11	3.00	1.00	3.25	0.96	3.00	1.29
Item 12	2.13	0.76	2.00	0.82	2.43	0.79
Item 13	3.35	0.65	4.25	0.50	2.86	0.69
Item 14	3.74	0.81	3.75	1.26	4.29	0.49
Item 15	4.04	0.88	4.50	0.58	4.00	0.00
Item 16	2.17	0.65	1.75	0.50	1.71	0.95
Item 17	3.74	0.86	3.00	0.82	3.86	1.07
Item 18	3.83	0.72	3.75	0.96	3.71	1.11
Item 19	3.87	0.81	3.75	0.96	3.43	1.27
Item 20	4.13	0.55	4.25	0.50	4.00	0.58
Item 21	4.17	0.49	4.75	0.50	4.00	0.82
Item 22	4.00	0.60	4.75	0.50	3.86	0.90
Item 23	3.52	0.90	3.75	0.96	2.86	0.90
Item 24	3.39	0.84	3.75	0.50	2.71	1.25
Item 25	3.35	0.83	3.25	0.96	2.71	1.25
Item 26	4.26	0.81	4.00	1.15	4.29	0.76
Item 27	2.65	1.27	2.50	1.29	2.43	1.13
Item 28	3.13	1.25	4.00	0.82	2.86	1.07
Item 29	3.70	1.11	4.25	0.96	3.57	0.79
Item 30	3.48	0.85	3.75	0.50	3.43	0.79
Item 31	3.39	0.78	3.25	0.50	3.00	0.82
Item 32	3.17	1.03	3.50	0.58	3.14	0.90
Item 33	3.26	0.86	4.25	0.96	2.86	0.90

Table 11 shows the means and the standard deviations for each item in the survey's first section for assessment literacy "knowledge" of English language professionals. The items in this category of the survey asked the participants to score the knowledge level of English language instructors from one to five meaning (1) not knowledgeable at all, (2) slightly knowledgeable, (3) moderately knowledgeable, (4) very knowledgeable, and (5) extremely knowledgeable. Each item refers to one of the seven dimensions in this category: *knowledge of theory*, *language pedagogy*, *principles and concepts*, *personal beliefs and attitudes*, *scores and decision making*, *local practices and impact* and *social values*. Except for few items with relatively lower means (to be explained in the following sections), the overall results suggest that English language teachers are *moderately knowledgeable* in assessment literacy as perceived by all groups. It can also be said that the perceptions of English language teachers based on Table 11 show both similarities and differences across different groups.

The item with the highest mean value as perceived by English language teachers themselves and academic leaders is item 26 which asks the assessment literacy knowledge regarding *the structure of language* ($M_1 = 4.26$, $M_3 = 4.29$). In a similar manner, assessment specialists also believed that the English language teachers are quite knowledgeable in the structure of language ($M_2 = 4.00$, $SD_2 = 1.15$). In addition, item 20 ($M_1 = 4.13$, $M_2 = 4.25$, $M_3 = 4.00$), item 21 ($M_1 = 4.17$, $M_2 = 4.75$, $M_3 = 4.00$) and item 22 ($M_1 = 4.00$, $M_2 = 4.75$, $M_3 = 3.86$) have high mean values as perceived by all groups, and all of the relevant items are categorized under the same dimension: *knowledge of theory*. In other words, as perceived by themselves, assessment specialists and academic leaders, English language teachers are *very knowledgeable* in this dimension. To illustrate, assessment specialists

thought that English language teachers are *very knowledgeable* in *how foreign/second languages are learned* ($M_2 = 4.75$) and in *how language is used in society* ($M_2 = 4.75$). Similarly, English language teachers ($M_1 = 4.17$, $M_1 = 4.00$) and academic leaders ($M_3 = 4.00$, $M_3 = 3.86$) also believed that English language teachers are quite knowledgeable in the relevant items.

In addition, academic leaders also believed that English language teachers are very knowledgeable in *how to prepare learners to take language assessments* (item 14) with a mean value of 4.29. This is another highest mean value as perceived by academic leaders, and the results also indicate that English language teachers ($M_1 = 3.74$) and assessment specialists ($M_2 = 3.75$) seem to agree.

The item with the lowest mean value perceived by all stakeholders is item 16. Thus, it can be suggested that English language teachers are *slightly knowledgeable* in *how assessments can be used to enforce social policies (e.g., immigration)* as perceived by all stakeholders ($M_1 = 2.17$, $M_2 = 1.75$, $M_3 = 1.71$). In a similar manner, item 12 ($M_1 = 2.13$, $M_2 = 2.00$, $M_3 = 2.43$) and item 27 ($M_1 = 2.65$, $M_2 = 2.50$, $M_3 = 2.43$) have lower means as perceived by all groups. These two items are categorized under the same dimension *local practices*; therefore, the results suggest that English language teachers are perceived as *slightly knowledgeable* in this dimension. On the other hand, the standard deviations for item 27 (the relevant legal regulations for assessment in Turkish preparatory schools) by all groups are over 1 ($SD_1 = 1.27$, $SD_2 = 1.29$, $SD_3 = 1.13$), which implies that there was no full agreement among the participants.

On the other hand, there were a few items which reflected noticeable differences across the three groups' perceptions. To illustrate, item 33 focused on *how one's own knowledge of language assessment might be further developed*. While

English language teachers and assessment specialists had means *above average* ($M_1 = 3.26$, $M_2 = 4.25$), academic leaders had lower means ($M_3 = 2.86$), which shows that the perceptions are different across groups.

Table 12 below presents the means and the standard deviations for each item in the assessment literacy “skills” section of the survey. The participants scored the skills of English language professionals from one to five meaning (1) not skilled at all, (2) slightly skilled, (3) moderately skilled, (4) very skilled, and (5) extremely skilled.

Table 12

Means and Standard Deviations for Assessment Literacy Levels: Skills Items

Items in the survey	Perceived by themselves (n = 23)		Perceived by assessment specialists (n=4)		Perceived by academic leaders (n=7)	
	M_1	SD_1	M_2	SD_2	M_3	SD_3
Item 34	2.43	1.16	1.50	0.58	1.57	0.79
Item 35	2.57	1.12	2.00	0.82	2.43	1.51
Item 36	2.35	1.03	1.75	0.96	1.71	0.95
Item 37	2.70	0.97	2.00	0.00	2.29	0.76
Item 38	3.96	0.88	4.25	0.50	4.14	1.07
Item 39	3.61	1.20	4.00	1.41	3.43	1.51
Item 40	3.57	1.12	4.00	0.00	3.14	0.90
Item 41	2.61	0.99	2.25	0.96	1.71	0.76
Item 42	3.04	0.88	3.00	0.82	2.86	1.21
Item 43	3.09	0.90	3.25	0.96	2.57	0.98
Item 44	2.22	1.00	2.25	0.50	3.00	1.41
Item 45	1.91	0.79	2.00	0.82	2.00	0.82
Item 46	2.74	0.86	2.25	0.96	1.86	0.69
Item 47	2.52	0.95	2.00	0.82	1.86	1.07
Item 48	2.39	0.89	2.00	0.82	1.71	1.11
Item 49	2.43	0.95	3.25	1.26	2.14	1.07

Table 12 (cont'd)*Means and Standard Deviations for Assessment Literacy Levels: Skills Items*

Items in the survey	Perceived by themselves (n = 23)		Perceived by assessment specialists (n=4)		Perceived by academic leaders (n=7)	
	M_1	SD_1	M_2	SD_2	M_3	SD_3
Item 50	2.17	0.98	3.00	1.15	2.86	1.46
Item 51	2.22	0.85	2.50	1.29	2.29	1.11
Item 52	2.87	1.10	3.00	1.41	2.43	1.13
Item 53	2.22	0.80	2.25	1.50	2.14	0.69

Table 12 indicates the mean values and standard deviations for the items in *skills* category. Each item refers to one of the three dimensions in this category: *constructing language assessments, administering and scoring language assessments and evaluating language assessments*. Except for few items with relatively higher means (to be explained in the following sections), the overall results suggest that English language teachers are *slightly skilled* in assessment literacy as perceived by all groups. Furthermore, many of the items have relatively high standard deviations, which imply that the participants did not have full agreement on their perceptions of the assessment literacy skills of English language teachers.

It is clear in Table 12 that English language teachers perceive themselves as *very skilled* in *using rating scales (rubrics) to score speaking or writing performances* (item 38) with the highest mean value in this section of the survey ($M_1 = 3.96$). Assessment specialists agree with English language professionals ($M_2 = 4.25$) on this item. In addition, item 42 and item 43 both of which are related to *the constructing language assessments* category have relatively high mean values as perceived by all groups. The mean values for item 42 are 3.04, 3.00, 2.86 and the

mean values for item 43 are 3.09, 3.25 and 2.57 as perceived by the teachers, assessment specialists and academic leaders respectively. However, the relatively high standard deviations for item 42 ($SD_1 = 0.88$, $SD_2 = 0.82$, $SD_3 = 1.23$) and item 43 ($SD_1 = 0.90$, $SD_2 = 0.96$, $SD_3 = 0.98$) infer that the participants did not have full agreement in their perceptions of these items.

The item with the lowest mean value as perceived by English language teachers themselves is item 45: *training others to write good quality items (questions) or tasks for language assessments* ($M_1 = 1.91$). Perceptions of assessment specialists ($M_2 = 2.00$) and academic leaders ($M_3 = 2.00$) were similar to that of English language teachers. Therefore, it can be suggested that teachers are *not skilled* in training others to write good quality items (questions) or tasks for language assessments, as perceived by all stakeholders. In addition, item 36 ($M_1 = 2.35$, $M_2 = 1.75$, $M_3 = 1.71$) and item 48 ($M_1 = 2.39$, $M_2 = 2.00$, $M_3 = 1.71$) have low mean values as perceived by all stakeholders, and both items are categorized under the same dimension, *evaluating language assessments*. However, the high standard deviations for item 36 ($SD_1 = 1.03$, $SD_2 = 0.96$, $SD_3 = 0.95$) and item 48 ($SD_1 = 0.89$, $SD_2 = 0.82$, $SD_3 = 1.11$) by all groups suggest that there was variation among the responses of the participants.

Furthermore, the item with the lowest mean value as perceived by both assessment specialists and academic leaders was item 34: *using statistics to analyze the difficulty of individual items (questions) or tasks* ($M_2 = 1.50$, $M_3 = 1.57$). In other words, assessment specialists and academic leaders believed that English language teachers are *not skilled* in using statistics to analyze the difficulty of individual items (questions) or tasks. Although English language teachers perceived themselves as

more skilled in this specific item ($M_1 = 2.43$), the standard deviation is relatively higher ($SD_1 = 1.16$), which implies a disagreement among participants.

When it comes to the item with the highest mean perceived by academic leaders, item 39 can be referred to ($M_3 = 3.43$). That is to say, academic leaders perceived English language teachers as quite skilled in *scoring closed-response questions (e.g. Multiple Choice Questions)*. The perceptions of English language teachers as well as assessment specialists were similar as the means were 3.61 and 4.00 in order. The survey results indicate that English language teachers are very skilled in scoring closed-response questions; but the high standard deviations in all stakeholders ($SD_1 = 1.20$, $SD_2 = 1.41$, $SD_3 = 1.51$) imply that the scores were not similar to each other.

In short, the survey results evidence that English language teachers are *more knowledgeable than skilled* in assessment literacy as perceived by themselves, assessment specialists and academic leaders. The dimension with the highest mean value (knowledge of theory) is categorized under the assessment literacy *knowledge* category. Seemingly, the dimension with the lowest mean (evaluating language assessments) is listed under the assessment literacy *skills* category.

To support the quantitative results from the survey and validate the findings, the qualitative results from the interview data will be presented in the following section.

Interview Results

In addition to the quantitative data from the assessment literacy survey, volunteer English language teachers were interviewed and asked to further elaborate on their assessment literacy levels. The transcribed data from the open-ended questions in the interview were analyzed through thematic content analysis in the

light of the categories and codes determined for each assessment literacy dimension within the theoretical framework (Taylor, 2013). In that respect, the perceived assessment literacy level of English language teachers will be further discussed with references to supporting and/or conflicting findings obtained from the survey results.

Findings from the interviews regarding assessment literacy levels of English language teachers are overall in agreement with the survey results. The qualitative data also suggested that the overall assessment literacy level of English language teachers is *around average*. During the interviews, quite many teachers stated that they realized their lack of knowledge and skills in assessment while filling out the survey used in this case study. For example, Teacher 9 expressed that the survey made her realize that she does not know much about assessment. Likewise, Teacher 6 said:

When answering the (survey) questions, I thought to myself that my assessment literacy level is not high. And I thought there were some things I thought that I should know, and I did not, as they seem important. It actually led me to think that maybe I should learn a little more about the assessment in general.

There were also many other teachers who clearly said that their assessment literacy level is mediocre. For example, Teacher 5, Teacher 1 and Teacher 7 stated that their assessment literacy is around average. They argued that although they took some courses back in their undergraduate and graduate studies, they do not feel assessment literate. Teacher 2 also said that she is not very good at assessment because although she took some graduate courses, she has no practical experience. Another teacher (Teacher 10) exemplified her mediocre level of assessment literacy as follows:

For example, summative assessment design, formative assessment design; I can explain the assessment design, but when you say let's go into the details, I stop.

Similarly, Teacher 7 said that she would not feel competent enough for certain assessment skills:

My assessment literacy level? I guess I would say average, maybe I feel like I would need more specific training by our testing unit. If I were to write questions for our assessments, I would need extra training about, you know, what exactly are we looking for? What kind of questions?

Therefore, it can be said that the interview results supported the survey results for the overall assessment literacy levels of English language teachers as both data sources affirmed that English language teachers are *moderately knowledgeable and slightly skilled* in assessment literacy.

To report the findings in more detail, firstly, the knowledge related dimensions are covered in the following sections.

Knowledge of Theory

The qualitative findings from the interviews regarding the teachers' knowledge of theory are also in agreement with the survey results to some extent. Although the surveys concluded that English teachers are perceived as very knowledgeable in the knowledge of theory dimension, there was not as much relevant data from the interviews. Yet, almost half of the teachers talked about their theoretical knowledge including the structure of English and English language learning during the interviews. One of the participants, Teacher 1, stated that for a comprehensive assessment, all skills that are in the English language must be tested: listening, reading, speaking and writing. Another instructor Teacher 9 said,

“Language is actually more like sports, like playing the guitar. The more practice, the better. There can be no test for this when you actually look at it”.

Language Pedagogy

Almost all participants articulated their knowledge regarding the language pedagogy dimension in the interviews. Thus, in line with the quantitative results, the qualitative data also confirm that English language teachers at the case institution are *very knowledgeable* in language pedagogy for assessment literacy. The most common concepts mentioned by the participants were washback, student achievement, feedback, and in-class assessment practices while they expressed their assessment literacy levels. Almost each participant gave examples of how they use assessment in class. They also shared a great number of experiences specifically related to the washback effect, which will be discussed in detail for the second research question. Teacher 1 said:

I would say that it (assessment) should be a gauge of how they're progressing, and they should be aware of improvements that they're making ideally. And that should have a knock-on effect as a form of motivation to encourage them to do even more.

Another teacher (Teacher 8) shared his perceptions on student achievement as follows:

A teacher's realization of how far his student has progressed and whether he will be successful in a classroom is also related to his own internal assessment. For example, the teacher can internally realize that students who look very good can fail. Why is that? He understands the sentence he speaks, understands the answer he gives, and understands his situation in the classroom.

Most of the participants also mentioned that they know how to use assessments to motivate students. Teacher 7 stated that being aware of assessment types and conveying this information to students to make them feel safe are among his strong features as a teacher. Another participant (Teacher 9) also touched upon assessment and student motivation by stating:

The value of the assessment is unknown. It is unknown to the student. For example, I always try to explain to students how important and valuable this is, in fact, and it is very useful for them.

As a result, interview results suggest that English language teachers are both knowledgeable and experienced in language pedagogy with regard to assessment literacy. Therefore, the qualitative findings for the dimension are parallel with the quantitative findings.

Principles and Concepts

For this dimension, more than half of the participants shared their knowledge and perceptions on the concepts of assessment *validity* and *reliability*. Most of them stated that they know what these fundamental principles of assessment mean. To illustrate, Teacher 1 said:

Assessment should assess what has been taught rather than things that have not been taught; so, for example, if they've not been taught how to summarize something, that should not be tested. All about validity.

Another instructor (Teacher 10) also shared his knowledge on standardization as follows:

Subjective testing could have a knock-on effect in that students might feel that you haven't treated them fairly. And the other classes have got higher points or higher grades, whereas you have given them lower ones, but if the

criteria are in place and good criteria and very clear criteria, and everybody abides by the criteria, it should be a fairly standard and reliable kind of assessment.

The qualitative results for this dimension are in line with the survey results to a great extent. Both English language teachers and assessment specialists thought that the teachers have above average knowledge of principles and concepts of assessments although academic leaders perceived them as moderately knowledgeable according to the survey results. The relevant comments presented above by the teachers support this quantitative finding, mainly perceived by themselves and by the assessment specialists.

Impact of Social Values

There were only two relevant reflections of assessment literacy level based on the impact of social values in the interviews. Both of the teachers discussed how *social values can influence language assessment design and use* (item 23 in the survey) during their interviews. Teacher 1 said:

I like to see students learn English because they want to, in general, not because of an exam, but in our students, we see kind of both types of students. So, this is because of the culture here. I mean I haven't lived in any other country besides the States and Turkey, but in Turkey there seems to be more of a focus or more pressure on assessment, formal assessment. And about this institution, that might be because you know, we have a department head who is an expert in assessment.

On the other hand, although English language professionals perceived the teachers as moderately/very knowledgeable in *how to recognize when an assessment is being used inappropriately* (item 13 in the survey) according to the survey, there

was no meaningful data to support this finding coming from the interviews. This lack of supportive data seems to conflict with the results of the survey which found English language teachers moderately knowledgeable.

Local Practices

This dimension mainly focused on the following three concepts:

Accreditation, Turkish assessment traditions and training others to increase assessment literacy level. Although the survey results suggest that English language teachers are moderately knowledgeable in these concepts of local practices, the interview yielded relevant results only on Turkish assessment traditions. That is to say, a few teachers shared their knowledge merely related to Turkish assessment traditions. Most of the perceptions shared during the interviews included negative criticism of Turkish assessment practices. For example, Teacher 9 said that assessment practices in Turkey seem really rigid in language education. Similarly, Teacher 5 made the following comment:

For the English language, the role of assessment in Turkey sounds like something problematic. In Turkey, for example, what we call YDS (a national foreign language test) does not measure whether they can use English in practice, it is exactly the same for students. no, but it seems like he can get very high marks from YDS.

Personal Beliefs and Attitudes

This dimension focused on English language professionals' knowledge in two main categories: *Awareness of personal beliefs and attitudes towards assessment including any conflicts with other groups and the ways to improve one's assessment literacy.* According to the survey results, teachers' assessment literacy level for this dimension is above average as perceived by all stakeholders. With a similar pattern,

there were countless comments on both topics during the interviews, which showed that they are in fact perceiving themselves as more aware of their own beliefs and attitudes towards assessment than what the survey results suggested. For instance, Teacher 8 clearly gave an overall understanding of his own belief on assessment as:

A teacher should be able to understand what can be assessed and how it can be assessed. And what can we do with the results of the assessment? And that's like an important, very important part of teaching overall in general.

Another teacher, Teacher 7, explained how her beliefs towards assessment conflict with other groups as follows:

Sometimes I feel a little bit unsure about how my classes fit with assessment, because personally, I feel more focused on teaching skills than teaching. And you know, readiness, is a certain part of the language, but maybe I could learn more about assessment or its effectiveness.

There could be one obvious and meaningful reason for the qualitative data to demonstrate higher awareness of personal beliefs and attitudes towards language assessment than do the quantitative data findings. The interview questions included specific questions on teachers' beliefs and attitudes towards assessments and a separate question on how to increase one's assessment literacy. Furthermore, all the participants shared a great deal of knowledge on how teachers can grow assessment literacy. Each participant suggested a variety of ways to improve assessment literacy ranging from online platforms to workshops. The detailed interpretations of this dimension as well as common training preferences will be discussed in the results section of the second research question.

Scores and Decision Making

Although the survey results suggest that English language professionals are quite knowledgeable in *scores and decision-making* dimension as perceived by themselves and by academic leaders, the interviews did not provide supportive data as none of the teachers talked about scores and decision-making regarding assessment literacy during the interviews.

The findings for the *skills* related dimensions in assessment literacy will be reported in the following section.

Constructing Language Assessments

This dimension mainly focused on English language professionals' skills in *constructing items, tasks, rubrics and test specifications*. The survey results showed that English language teachers are slightly skilled as perceived by the academic leaders and moderately skilled as perceived by assessment specialists and the teachers themselves. The interview results partially confirmed this as most teachers' skills pointed out for this dimension were limited to *constructing items for in-class practice*. Teacher 10, Teacher 5, Teacher 7 and Teacher 9 stated that they can write multiple choice type of questions while adapting class materials. Teacher 10 exemplifies how he does this for listening questions as follows:

Sometimes the listening is very complex, and sometimes the questions of the book are not very targeted, then I am writing just a couple of multiple-choice questions. I immediately project onto the screen and ask them to respond.

This is something with TOEFL.

Similarly, half of the teachers confirmed that they would need further training to write quality items. For instance, Teacher 7 stated that if she were to write questions for assessments, she would need extra training about the content and type

of the questions. In addition, one of the teachers (Teacher 5) and a former member of the Curriculum Development Unit of the case institution acknowledged her lack of skills in writing quality items:

I have actually prepared a lot of materials in my curriculum but not exactly an assessment. For example, I prepared extra material last week, but when I applied it in class, I didn't like it at all. To be honest, I see my deficiency and I'm not going over it right now.

None of the teachers discussed their skills in designing rubrics or preparing test specifications during the interviews, possibly because they had no relevant experience.

Administering and Scoring Language Assessments

Scoring closed-ended questions and *grading productive skills* were the main categories of this dimension. The survey results displayed that English language teachers are moderately skilled in this dimension. However, two participants talked about their skills in grading speaking and writing assessments of the students during the interviews. On the other hand, none of them touched upon scoring closed-ended questions, possibly because the closed-ended questions of the exams at the institution are scored by assessment specialists and shared with the teachers.

Evaluating Language Assessments

This dimension focused on teachers' *skills in statistics* for evaluating language assessments as well as *cut-scores*. It was found out in the survey that English language teachers are perceived as slightly skilled in evaluating language assessments. However, there were no relevant qualitative data obtained from the interviews as none of the teachers touched upon these skills during the interviews.

English Language Professionals' Perception of the Relationship Between Assessment Literacy and Teaching Quality

The second research question was explored through the qualitative data from the interviews. In the thematic content analysis process, several of the dimensions which were related to teaching quality within Taylor's framework (2013) were discussed by the participants. Although there were some variations among the views of different groups, participants' responses were mostly gathered under *language pedagogy, personal beliefs and attitudes, constructing language assessments* dimensions.

English Language Teachers' Perception

Through the means of the interview questions, English language teachers shared their perceptions of the relationship between assessment literacy and teaching quality. Overall, most of the teachers believed that all English language teachers should be assessment literate to a certain extent in order to foster learning and teaching in the classroom. Besides, they mentioned their need to improve their assessment literacy to be more effective as a teacher. The detailed findings will be presented for each dimension of assessment literacy along with relevant comments made by the teachers in the following section.

Firstly, the knowledge related dimensions of assessment literacy are covered in the next section.

Knowledge of Theory

Not many of the teachers mentioned the relationship between knowledge of theory and teaching quality. However, some stated that different aspects of a language such as knowledge and skills should be measured for a comprehensive assessment.

Language Pedagogy

Almost all of the participants referred to language pedagogy while they explained the relationship between assessment literacy and teaching quality. Therefore, it can be said that language pedagogy is the most relevant dimension of assessment literacy as perceived by English language teachers. The most frequently emerging themes were *washback*, *feedback* and *in-class assessment practices*, while they expressed their perceptions of the relationship between assessment literacy and teaching quality.

Washback effect was the most commonly appearing category in the interviews. As the nature of the concept itself is both complicated and comprehensive, there were a great number of related comments made by the participants. Overall, qualitative findings show that English language teachers believe assessment practices at the institution have a huge impact on their teaching. Most of the participants implied that the effect was positive for their teaching quality. One of the participants (Teacher 2) emphasized the importance of familiarity with assessments to shape instruction accordingly as follows:

If I know what the assessment tool is, then I feel like I have more clarity and more alignment with the assessment tool and with the curriculum. I think it shapes my priorities for the class. I feel like I can prepare students more for that assessment.

A lot of the teachers also stated that the students feel more motivated with teachers who are knowledgeable about the assessment practices at the institution. For example, Teacher 7 stated:

I think it (awareness of question types) is helpful in this way, because students think they are safe, they always need a rule. For example, they want

me to see something and mark it immediately. And when I can provide it somehow, they feel safe, they feel more confident.

Another teacher (Teacher 6) stated that the exam system at the institution has shaped her reading classes a lot. Because there are certain types of questions in the exams, she gives those test questions first instead of the reading questions in the book. The important thing seems to her that the student is trying to gain that test practice, and it causes students to focus better. Teacher 4 also believed that teachers should shape their teaching based on assessment results. Other teachers believed the main role of assessment is to provide feedback for the instruction. One teacher, Teacher 6, said:

Assessment is important because I can modify my teaching according to the results. Just like I said, when you see the problems in a student's learning, it means a difference between the goals and the student's actual learning. You can actually modify your teaching to reach the goals.

With a similar perspective, Teacher 3 said that assessment helps a teacher to shape teaching.

The role of assessments for me is to know where the student is at. And it gives me a few directions, whether they grasp what was being taught. And if they did, it tells me where I need to change my teaching or what I need to do to supplement it where they are, and then to make changes to the instruction.

Almost all participants said that it is important to refer to assessment to progress their students' achievement. One of the teachers (Teacher 3) said that the role of assessment is to know where her students are at and to understand whether they grasp what is being taught. Similarly, Teacher 8 believed that thanks to assessment, one can understand in which skills a class has progressed as the lessons

go on. Another teacher (Teacher 8) shared her perceptions on positive washback by exemplifying how she uses assessment results of productive skills to improve student learning in class as follows:

For example, you are looking, let's say especially in writing, over the essays that the child writes or the speeches he makes, you can say that this child is missing. You're trying to direct him more there. Or it happens more in grammar because we are more into grammar. You know, when you look at the progress of the child's grades or the things he does in the classroom, then you say that 'you come at the office time, and we will study them with you again'.

On the other hand, only a couple of teachers perceived washback as a negative effect on their teaching quality. They said that when they shape their teaching according to the assessment practices at the institution, it conflicts with their teaching philosophies. Besides, this prevented the institution and the teachers from using alternative assessment methods although they wished for it. One teacher (Teacher 5) discussed the way students are motivated by exam-oriented instruction although she does not like it as follows:

For students, exam results are very important, after all, this is a corporate culture. Actually, I personally don't like it, but it affects the students in that way and motivates them, and at the end of the day I become one of the test teachers because their motivation is low.

TOEFL ITP was very frequently mentioned by the teachers, especially with reference to the fact that it consists of items only for receptive skills. They talked about the negative impacts of this on their teaching such as when the students tend to be reluctant in engaging in productive skills. They alleged that certain assessment

practices at the case institution including TOEFL ITP turned them into a ‘test teacher’. One teacher (Teacher 9) explained how the proficiency test affects students’ engagement in class in the following comment:

In this school, too, at the end of the year, the students are expected to pass a test like TOEFL. It's like a process which can negatively affect my courses as follows; When I try to speak, I encounter resistance. "Sir, explain grammar," they say. When I give a grammar test or a vocab test, or reading tests given by the material office, everyone agrees.

Feedback was another concept touched upon a considerate number of times by the English language teachers during the interviews. Most of the participants clearly stated that feedback based on assessments is a valuable component of the curriculum. They often referred to the feedback hour allocated for each written exam at the case institution, and said that it is helpful for students to see their papers soon after they get the results because they have a chance to go over the questions along with a teacher. One teacher (Teacher 6) also asserted that feedback is essential to improve teaching quality in the following way:

Especially when it comes to teaching, when it comes to giving feedback in class, you know managing the further steps after the evaluation is important. So, when you can provide proper feedback and when you can understand the results of an assessment, it will help you a lot in the process of teaching and the process of improving students’ understanding.

Furthermore, a few teachers (Teacher 4, Teacher 3, Teacher 2) felt that the way they need more time to provide feedback to their students. To illustrate, Teacher 4 said:

It takes like five to 10 minutes to give feedback to one student, especially in lower-level classes. So that's why the timing might be a problem when it comes to feedback.

One of the teachers (Teacher 10) also said that as teachers of productive skills, they spend a great deal of time checking students' papers to provide feedback:

After all, we test and give feedback every week, and I think giving feedback is very important and good feedback. Sometimes it is not easy in writing because it is not marked while being read. So as a writing teacher again... It takes a while. You spend an hour, a full hour, reading the writing one by one, but I still think it is useful.

Lastly, almost all of the teachers agreed in the interviews that using *in-class assessment practices* improved teaching quality overall. One of the teachers (Teacher 5) exemplified how her in-class assessment practices and student concentration go hand in hand:

I find (assessment) materials that I like and use them in different ways. I turn a shared material into a format I want exactly, according to the needs of the class. Because when I see that the student concentrates better on it. I'm preparing little things on the board at work. I have my own assessment tools.

Most pronounced in-class assessment tools include peer-assessment, self-assessment and online assessment. For instance, Teacher 7 stated that she uses peer assessment where students mark each other's writing papers. Similarly, Teacher 4 believed that peer-assessment practices help students learn from each other:

I try to encourage the students to use peer-assessment, but it's really informal, like discussing these questions with your partner and then as they're

discussing, maybe they can understand, Okay, I'm good at talking about this topic, but not this topic, you know?

Besides, one teacher (Teacher 3) refers to her self-assessment techniques to observe learning as follows:

I'd really like to do formative assessment checking, to see where the students are. And also a little bit of where self-reflection fits under assessment, such as asking students, for example, we listened to something and I asked them, especially for CR level again. How was the listening? What percent did you get?

As for online assessment, Kahoot, Quizlet and Quizzes were the most frequent online tools discussed in the interviews. It should also be noted that teachers mostly said they use the online assessment tools provided by the Curriculum Development Unit of the institution. Teacher 2, Teacher 3, Teacher 4, Teacher 6 and Teacher 10 stated that online assessment tools are valuable because we live in the era of technology which can provide us with many conveniences such as instant results in online assessments. The teachers often stated that they use the online assessment to provide feedback for students. Teacher 6 gave an example of using online assessment tools to be informed about student levels as follows:

Well, actually I can use the live Quizzes we do in class or that kind of stuff.

When we use them in class, it actually shows in real time what might be the problems with the students. And you can actually, after that quiz is finished in class, give feedback to students about the problems they have.

Principles and Concepts

The majority of the teachers stated their opinions related to the validity and reliability of assessments when they described the practices at the institution. Most of

them said that it is a good idea to have *a separate unit* who are experts in assessment for valid and reliable assessment. They believed that it requires special expertise to manage the assessment processes in a valid and a reliable way. One of the teachers (Teacher 10) shared her perceptions on the way that assessment specialists ensure content validity in the following comment:

I find it (the assessment system here) very professional, and I think we ask for most or even everything we teach. We are speaking in class, we have assessment; we are doing writing and we have assessment, we do reading and grammar, they come out in the exams. I think they are prepared in such a professional way with high quality.

Many other teachers also believed that for higher quality teaching, a specific group should be involved in the design of assessment practices. They frequently mentioned the standardization meetings where there is an assessment specialist to lead the marking sessions. They also said that grading students' writing as pairs of teachers and having an assessment specialist to consult when there is an issue is an effective strategy to increase reliability. One teacher (Teacher 8) who has worked at the case institution for seventeen years and is a former academic leader shared her perceptions of having a separate unit to manage assessment processes by referring to specific past experiences:

If we would like to share everything about assessments, then our reliability will collapse. So, we crash. I think it's good to have an office. In fact, if we were democratic, we'd have such a pool of questions, but we tried it in the past. There were such big problems that we passed out while editing it all. So experienced professors who have a good knowledge of the job should do it; an office is fine.

On the other hand, there were some concerns raised by the teachers about the *portfolio* practices at the institution in terms of validity and reliability. About half of the teachers believed that they do not serve their best under the given circumstances. The underlying reason for this perception of teachers was found to be the design of the portfolio tasks as well as using the same tasks across different levels/trimesters. Teachers believed this negatively affect teaching and learning. One teacher (Teacher 8) stated her opinions about the portfolio design as follows:

I think portfolios are not very successful. In other words, I look at it from the students' point of view, that is, they make those portfolios with a bit of joint work in the previous semester and the next semester.

Another teacher (Teacher 5) also exemplified an issue related to the validity and reliability of the portfolio tasks in the following way:

I am a professor who thinks that the portfolio should continue. I can recognize the student who cheated on his writing assignments, as he is doing perfect homework. He either used translation or got help from someone else. But since I couldn't prove it, I have to give it high marks. There is a portfolio supposedly, but I don't think I can get the result.

In addition, *TOEFL ITP* was pronounced by a few teachers as an influence on the validity and reliability of the assessment practices at the institution. Most stated that there is a tendency to construct language assessments similar to the ones in *TOEFL ITP*, mostly multiple choice. They said assessment specialists mostly include multiple-choice questions in the exams because *TOEFL ITP* is a multiple-choice test. Teachers believed that the students tend to memorize the grammar rules or commonly asked topics of the exam. They said this negative washback decreased the

validity and reliability of the assessment practices as the proficiency exam does not include a section for productive skills.

I think the thing is because of TOEFL as an exit exam, you know, that really conflicts with students who want to improve their speaking skills. Speaking skills is something that shows that they are able to know what to do in English. So, my main concern is whether the students, who are able to take the exit exam and leave the department, know how to do it.

Impact of Social Values

Only one of the teachers, who was a non-Turkish participant (Teacher 7) discussed the effect of social values on her teaching quality. As a native speaker and teacher of English, she perceived Turkish assessment practices to be very strict in the following comment she made:

I mean, sometimes I feel like not only in our school, but in Turkey assessment seems really rigid in language education. And sometimes I wonder if it needs to be that rigid or that important. But again, I feel like I'm kind of outside coming in and so more like an observer and just going along with how the system works.

Local Practices

None of the participants talked about accreditation and/or training others on assessment literacy during the interviews. However, a few of the participants claimed that the effect of Turkish assessment traditions is visible at the case institution. Teacher 1, Teacher 4, Teacher 5, Teacher 8 and Teacher 10 stated that there is too much emphasis on assessment in Turkey and this negatively affects their teaching quality. Teacher 5 also believed that most Turkish students like the ones at the

institution improve most in grammar, so the teachers are also spending more time on grammar.

Personal Beliefs and Attitudes

The first focus of this dimension was teachers' awareness of their personal attitudes towards assessment and how it sometimes conflicted with the others at the institution. To start with, most English teachers stated that filling out the survey was difficult because they did not know about or experience many of the items. However, they agreed that assessment literacy is important. One teacher, Teacher 6, said that a teacher should be able to understand what can be assessed, how it can be assessed, and what can be done with the results of the assessment, which he believed is a very important part of teaching overall. Another teacher (Teacher 7) summarized her personal attitudes towards assessment literacy in the following way:

I think it (assessment) is very important. Because it is related to both this lecturer and the fact that the teacher feels confident and comfortable. It's about adjusting the syllabus accordingly. And in return for this, we ultimately test the learning levels of the students. And we see the reflection of that again. So actually, I think it would be more beneficial if the instructor knows this, if he knows the way of testing or how to test.

In addition, there were a few common comments saying that teachers' personal attitudes towards assessment sometimes conflict with the assessment practices at the institution, partly due to the proficiency exam. Many believed that people pay too much attention to the assessment practices especially the proficiency exam. As a result, students as well as other stakeholders tend to focus more on multiple-choice type of questions. One of the teachers (Teacher 5) who is a former

member of the Curriculum Development Unit also thought that the quality of the assessments and teaching could be better without such a proficiency exam:

If it weren't for TOEFL, we might have a little more space. More ideal exams would be prepared. It would not be possible to say that it will come out in this exam. It's not unique to me here. I think all teachers are like that.

The second focus of the dimension, *training ways to improve assessment literacy*, was explored through the long list of strategies suggested by the participants. Teachers talked about their training experiences as well as preferences during the interviews. The methods shared by the teachers were having a degree in assessment, review of relevant literature, online courses, observing other institutions and gaining first-hand experience in assessment processes such as constructing items. In this regard, more than half of the participants mentioned the Peer-Learning sessions offered by the assessment specialists at the institution. Those who attended shared their positive feedback and those who could not make it wondered if such training would be given again. One teacher (Teacher 1) made the following comment:

I think you (assessment specialists) have in the past given presentations, peer learning, So that's quite useful as the point of doing that is to open a window, so to speak, to allow people to peer in and see what it's really like being in an assessment and measurement unit.

Scores and Decision Making

None of the teachers established a relationship between teaching quality and scores and decision-making dimension. This could be because cut scores and institution wise decision making are managed by the university administration and academic leaders, not by the teachers at the case institution.

The following sections will present the findings for the *skills* related dimensions in assessment literacy.

Constructing Language Assessments

The perceptions of the English teachers for this dimension of assessment literacy skills mostly gathered around constructing test items. A few of the teachers asserted that their skills in writing items especially multiple-choice questions contribute to teaching quality. They said that by knowing how to create test items, teachers can adapt the materials in the coursebooks when necessary. To illustrate, Teacher 10 said that she writes multiple-choice type of listening questions before her students listen to an audio in class, because the exercises in the book are either too complex or not targeted. Another teacher (Teacher 3) said that a teacher needs competence in constructing language test items to support the curriculum as follows:

There are pros and cons of having a separate unit. One of the pros is you guys have specialties in it and you know how to do it correctly. And the assessment is so important. You know, we need people who are experts in it. But on the other hand, if I want to help the curriculum and make some materials, like worksheets, I think I need to know how to create those types of questions, multiple choice questions.

However, some other participants did not establish a relationship between a teacher's skills in constructing language assessment and teaching quality. For one thing, they said that teachers are not the ones creating tests, so they do not have much experience. In addition, most believed that they do not need to be very skilled in item writing as there are assessment specialists doing this for them. For instance, Teacher 1 said:

How would my skills in constructing language tests affect my teaching quality? I got to think about that. Well, I'm not the one creating the tests, or the portfolio assessment tools. I'm the one implementing those.

Similarly, Teacher 2 stated that she believed teachers do not need to come to a level where they can create tests since there is a separate group of specialists doing the job for them.

Administering and Scoring Language Assessments

The few perceptions shared for this dimension were limited to grading productive skills. One of the teachers, Teacher 5, stated that rubrics and criteria should be well-defined for fair assessment of speaking and writing. In addition, Teacher 4 said that an overall understanding of rubrics for productive skills is beneficial to teachers and students. She also mentioned the importance of standardization of scoring productive skills by an assessment specialist to consult during the marking sessions.

Evaluating Language Assessments

None of the teachers touched upon their skills in using statistics to evaluate language assessments during the interviews. This may suggest that English language teachers did not perceive a relevant relationship between their skills in using statistics and their teaching quality, possibly because they are not skilled in this dimension.

To conclude, the majority of the English language teachers believed that a teacher's assessment literacy is related to their teaching quality to a certain extent. Almost all of them discussed their perceptions and experiences related to the *washback* category, while none mentioned *using statistics to evaluate language assessments* and *scores and decision making* during the interviews.

Assessment Specialists' Perceptions

In this section, first the perceived assessment literacy levels of assessment specialists based on the interview data will be presented dimension by dimension. After that, the perceptions of assessment specialists on the relationship between assessment literacy and teaching quality will be further explained again based on each dimension.

Perceived Assessment Literacy Level of Assessment Specialists

Perceived assessment literacy levels of assessment specialists were explored through the interviews conducted with four assessment specialists (AS). Each of the participants stated that since they work as assessment specialists at the case institution, they believe that they have a higher assessment literacy level when compared to the other teachers at the institution. They perceived themselves as *quite knowledgeable and skilled* in assessment overall. However, two out of the four assessment specialists specified their need for more theoretical knowledge in assessment. Assessment Specialist 3 stated that her theoretical knowledge in assessment is not that good, but she could say that her hands-on knowledge is good. Similarly, Assessment Specialist 4 shared her lack of knowledge in theory when she made the following comment:

Finding a class that would teach the basics of measurement and evaluation would help me because I just need the basics, the theories, the terms of it, because I know what I'm doing, but I don't know, for example, the name of what I'm doing or the statistics version of it. Maybe sometimes I don't know why I'm doing it. And the theory part would help me to do that.

To report the findings in detail, firstly, the *knowledge* related dimensions are presented in the following sections.

Knowledge of Theory. For this dimension, only one of the participants shared his perception on *the structure of English*. Assessment Specialist 1 stated that as assessment specialists, they try to assess the students' ability of speaking, listening and writing, and all kinds of skills and aspects of the language. He also noted that different skills require different assessment methods including alternative assessment.

The fact that most participants did not acknowledge theoretical knowledge during the interviews is in agreement with the perceptions of the two assessment specialists who said that they lack theoretical knowledge related to assessment although they feel competent in practice. Besides, the training preferences shared by the assessment specialists are mostly formal education such as pursuing a Ph.D. in assessment to improve their assessment literacy level. Therefore, qualitative findings suggest that assessment specialists do not perceive themselves as much knowledgeable in the knowledge of theory.

Language Pedagogy. All assessment specialists talked about their perceptions on *washback* during their interviews. Assessment Specialist 1 stated that he looks at the assessment results and thinks everybody should do that to mirror students' success, and adds that this depends on the student's motivation. Assessment Specialist 2 also made the following comment:

For example, after giving feedback, I write down some problematic areas or some questions like most of the students got wrong, and then we go over it if we have time. Most probably on Monday, not on Tuesday, right after the feedback. And so it's a way of seeing yourself from that point.

Similarly, Assessment Specialist 4 says that after checking the grades that students got from quizzes or midterms, she can easily say that they have some

problems in some topics or subjects, so she needs to revise it. Another assessment specialist (Assessment Specialist 3) highlights that as a teacher she knows how to teach her students what they are going to be facing in the exams, what kind of questions or types of assessments they will be seeing. She also said that some of her peer-assessment practices involve the use of rubrics used in the exams. She made the following comment:

I try to have my students mark each other's papers, for example, before the speaking quiz, I match them together as pairs. And then, I tell them to grade each other, according to the rubric.

Overall, qualitative findings suggest that the assessment specialists at the institution perceive themselves as *quite knowledgeable in language pedagogy* dimension of assessment literacy. Common concepts regarding washback and in-class assessments will be further discussed in the results of the second research question.

Principles and Concepts. There were not many results obtained from the interviews specific to the principles and concepts of assessment literacy. Only one participant, Assessment Specialist 4, stated that because she works as an assessment specialist, she knows what reliable and valid assessment is. As this dimension mainly focused on knowledge of *validity and reliability*, it can be said that some assessment specialists lacked theoretical knowledge of the principles and concepts as perceived by themselves.

Impact of Social Values. None of the assessment specialists discussed their knowledge of the impact of *social values* on assessment practices during the interviews.

Local Practices. None of the assessment specialists shared their knowledge regarding accreditation. This implies that they are not knowledgeable in the local assessment practices. However, two of the assessment specialists reflected on Turkish assessment practices, especially when they compared the case institution to the other preparatory schools. To exemplify, Assessment Specialist 2 said:

In preparatory schools in Turkey, in general, the role (of assessment) is to prepare students for departments, I guess. And if they have English medium classes or after university, if they're going to use it, it's going to be helpful for the students.

Personal Beliefs and Attitudes. All of the assessment specialists demonstrated awareness of *personal beliefs and attitudes towards assessment* including any *conflicts with other groups* and *the ways to improve one's assessment literacy*. Therefore, the results of the interviews suggest that assessment specialists are quite aware of their personal beliefs towards assessment. Assessment Specialist 1 and Assessment Specialist 2 believed that they tend to focus on assessment too much because of their role as assessment specialists. Assessment Specialist 1 explained how his personal beliefs conflicted with the current atmosphere as follows:

For this institution, I can say that we have an environment where students feel a little bit anxious, and they have stress about the assessment of their learning skills. So, also the teachers have it. I can see that the teachers are stressed about the assessment. So we can say that there's a negative pushback effect here. So we should maybe take the stress out of the classroom, but I don't know how.

Another assessment specialist also expressed her personal and conflicting beliefs towards the assessment schedule of the case institution by making the following comment:

With the assessment system here we have some problems, especially for the productive parts. For example, we have more than two exams or quizzes that assess students' writing skills, but when it comes to speaking, we have only a presentation. That's just five points. And then we have a speaking exam and just nine points. That means like 14 out of a hundred. I think that's not enough.

Overall, all assessment specialists shared their personal beliefs and attitudes towards assessment in different ways. In addition, various strategies to improve assessment literacy were suggested in the interviews. This shows that the assessment specialists at the case institution are quite knowledgeable about the ways of training in assessment literacy, and the common perceptions will be further discussed in the results of the second research question.

Scores and Decision Making. Most of the assessment specialists talked about scores and decision making during the interviews. Therefore, it can be suggested that assessment specialists at the case institution are very knowledgeable in score interpretation as perceived by themselves. They all said that they refer to students' scores to ease decision making processes both at the classroom and institutional levels. One of the assessment specialists, Assessment Specialist 2, made the following comment:

Of course, we are deciding if the students pass or fail. That's the most basic aim in that, but in class, the role of student assessment is to decide if a student

needs improvement, needs more support, or if he is good on his own, because he's going to use English in his life.

In the following sections, the findings for the *skills* related dimensions in assessment literacy will be covered.

Constructing Language Assessments. As for *constructing test items, creating rubrics and test specifications*, assessment specialists regarded themselves as quite skilled. Assessment Specialist 1, Assessment Specialist 2 and Assessment Specialist 4 said that they can create new assessments, construct test items and prepare test specifications. Assessment Specialist 4 gave an example of her skills in item writing:

Since I'm a member of the testing unit, I can easily form some questions. I remember one time that I was a substitute teacher in one of the classes and while they were reading, I wrote down some questions and they were really shocked. That was the first time they encountered questions in a written form provided by their main course teacher.

Although Assessment Specialist 3 also said that her hands-on skills are good in general, she did not specifically refer to each skill for this dimension. However, the overall findings from the interviews suggest that assessment specialists are quite skilled in constructing language assessments as perceived by themselves.

Administering and Scoring Language Assessments. None of the assessment specialists pronounced their skills in administering language assessments during the interviews.

Evaluating Language Assessments. Only one of the assessment specialists talked about her skills in *statistics* for evaluating language assessments. Assessment

Specialist 2 said that she is very skilled in item analysis. However, none of the assessment specialists acknowledged their skills in determining cut-scores.

Although there were not many perceptions on technical skills discussed by the assessment specialists in the interviews, it should be noted that all of them perceived themselves as *better in hands-on practices than in theory* of assessment when they were asked to evaluate their assessment literacy levels. The reason why they did not need to specify their technical skills could be that the researcher herself was an assessment specialist; and thus, they could have assumed that she already knew what is meant by ‘hands-on’ skills.

Assessment Specialists’ Perception of the Relationship Between Assessment Literacy and Teaching Quality

Except for one assessment specialist (Assessment Specialist 1), the other assessment specialists (Assessment Specialist 2, Assessment Specialist 3 and Assessment Specialist 4) alleged that their assessment literacy contributed to their teaching quality in various ways. They said that having experience in certain dimensions of assessment literacy helps them perform their teaching in a more effective way. The most common themes they referred to while explaining the positive relationship between assessment literacy and teaching quality were awareness of assessment types and regulations as well as some technical skills such as constructing test items.

The following sections will cover assessment specialists’ perception of the relationship between knowledge related dimension of assessment literacy and teaching quality.

Knowledge of Theory. For this dimension of assessment literacy, three assessment specialists (Assessment Specialist 1, Assessment Specialist 2 and

Assessment Specialist 3) discussed their perceptions on the relationship between the structure of the English language and teaching quality. Assessment Specialist 1 said the assessment methods at the institution cover all the skills in the English language, which empowers teaching and learning. He also mentioned that all aspects of English are measured through different tools such as portfolios. Assessment Specialist 3 also stated that assessment should include evaluation of life skills as the subject of matter is a language. On the other hand, Assessment Specialist 4 discussed some drawbacks of the current system regarding language skills. She believed that all skills should be given the same importance for quality assessment and teaching, which is not the case in her perception. She said:

With the assessment system here we have some problems, especially for the productive parts. For example, we have more than two exams or quizzes that assess students' writing skills, but when it comes to speaking, we have only a presentation. That's just five points. And then we have a speaking exam and just nine points. That means like 14 out of a hundred. I think that's not enough.

Language Pedagogy. Similar to the case with English language teachers, assessment specialists also discussed a great number of topics related to their language pedagogy and teaching quality. The most pronounced category by the assessment specialists was washback.

Three of the assessment specialists (Assessment Specialist 2, Assessment Specialist 3 and Assessment Specialist 4) perceived *washback* as a positive influence on their teaching. They believed that their literacy and experience in assessment affects their teaching in a positive way. Assessment Specialist 2 said:

Assessment shows me the way and leads me to the correct path. And I use the results of my assessments and classroom assessments to shape my teachings in the classroom.

In the same way, Assessment Specialist 4 gave an example of how she uses assessment results to shape her instruction:

For example, after giving feedback, I write down some problematic areas or some questions like most of the students got them wrong, and then we go over it if we have time. Most probably on Monday, not on Tuesday, right after the feedback. And so it's a way of seeing yourself from that point.

On the other hand, Assessment Specialist 1 perceived washback as a negative push toward teaching quality. He both asserted that assessment literacy would not necessarily affect one's teaching quality and claimed that the case institution is suffering from a negative washback effect. He added that this is the case with many other institutions as well with the following comment:

So I have a kind of tendency to emphasize the subject in such a way that they will be asked in the quizzes or in the midterms. I know that's not a good thing, but sometimes it affects teaching.

Different from the English language teachers, none of the assessment specialists recognized the effect of in-class assessments on their teaching quality. They did not prefer using online assessment tools such as Kahoot or others. With the same respect, none referred to self-assessment, which was frequently discussed in the interviews with the English language teachers. However, Assessment Specialist 1 said that he makes use of personal talks with his students to assess their achievement levels. Besides, Assessment Specialist 4 mentioned peer-assessment in the following way:

I try to have my students mark each other's papers, for example, before the speaking quiz, I try to match them together as pairs. And then, I tell them to grade each other, according to the rubric.

Principles and Concepts. Most assessment specialists noted the validity and reliability of the language tests used in a systematic and professional way at the case institution. They said that the assessment schedule is quite comprehensive and aiming at all skills in the English language.

However, there were some concerns shared about the validity and reliability of the productive skills, mainly speaking skill. Assessment Specialist 4, who is the most experienced assessment specialist at the institution, said that the design of the speaking exam has been the same over the past ten years and this negatively affects teaching and learning overall. Assessment Specialist 2 also claimed portfolio tasks need revising:

As for the portfolio, the design is good, but it becomes repetitive at some point. However, I love the stages in it, you know, they're making presentations and it's useful. But when you do it three times in a year, it loses the meaning of it.

Impact of Social Values. None of the assessment specialists touched upon how social values affected their assessment practices and teaching quality.

Local Practices. Similar to the dimension above, none of the assessment specialists talked about their insights about Turkish assessment traditions and its effect on their teaching quality.

Personal Beliefs and Attitudes. Half of the assessment specialists expressed their opinions on how their personal beliefs in relation to teaching quality conflict with the current practices. For instance, Assessment Specialist 1 said that the

workload at the institution prevents assessment specialists from varying the assessment tools.

The exams are just two parallel and it's difficult to include different question types. But there are also other issues which prevent us from doing that, you know, the difficulty of grading, the workload, and those things.

Similar to what several English language teachers said, some assessment specialists also called the name of the head of the department as putting *too much emphasis on assessment*, which creates an exam-oriented environment. They said that it creates stress on both students and teachers.

When it comes to *ways to improve assessment literacy*, except for Assessment Specialist 1 who is already pursuing a Ph.D. in measurement and evaluation, three assessment specialists shared a number of ways of training. Their areas for improvement changed, though. For example, Assessment Specialist 3 said that she would need to learn more about the theoretical background of assessment. However, Assessment Specialist 1 and Assessment Specialist 2 said they could improve in statistics to increase their assessment literacy level. All of the three mentioned *taking online courses, discovering what other institutions do, pursuing a Ph.D. in the field* to improve their assessment literacy. On the contrary, Assessment Specialist 1 believed that he does not need or prefer to improve more on assessment as this mindset bears a negative washback effect on his teaching quality.

Scores and Decision Making. Most assessment specialists mentioned that scores can be used as evidence to help decision-making processes. They mentioned the practice of putting classes in order of achievement based on assessment results. The results serve the institution as performance indicators of teachers. Assessment Specialist 4 also refers to this at institutional levels by saying:

I think I explained the subject clearly and I think my students understood it, but do you have any evidence and that's the evidence we have, the points they got, of course not the only evidence, but one which is also useful for the institution. It always puts classes in order to check their averages and then puts down them in a graph. I think it's really important for him (the department head).

However, Assessment Specialist 2 had some doubts about the way the institution interprets assessment results. She mentions how the institution calculates the achievement levels of classes taught by each teacher and then shares the results as evidence in performance appraisal meetings. Although she believed that the results can be regarded as evidence for student learning, she questions the way the institution uses it as a factor in evaluating teachers' performance.

The following sections will present the findings for the *skills* related dimensions in assessment literacy.

Constructing Language Assessments. Half of the assessment specialists pronounced no relationship between their skills in constructing language assessments and teaching quality. The other two assessment specialists stated that their skills in constructing language assessments such as multiple-choice questions increase teaching quality. Assessment Specialist 4 exemplifies a time when she made use of her item writing skills to diversify her teaching practices in a class in the following comment:

Because of the fact that I am a member of the testing, I can create questions in class while students are listening to or reading something. Otherwise, I would just distribute the handout or the worksheet given by the curriculum. So, these skills affected me in a good way.

Similarly, Assessment Specialist 3 mentioned that there is a little time gap between when she started working as a teacher and as an assessment specialist. Therefore, she improved her teaching skills through the skills she has got in item writing.

Because when I started here in testing, I didn't know how to be a teacher either. It was also my first experience as a teacher. So, it's helped me to better teach English because I tried to understand what types of questions I might ask my students.

On the whole, half of the assessment specialists perceived their skills in constructing language tests add to their teaching, while the other half did not recognize this during the interviews. In addition, three of the four assessment specialists stated that they should involve other teachers in the process of item writing such as in the form of getting feedback for test specifications. Although they were concerned about the secrecy issue, they mostly believed that all teachers should be more involved in constructing language assessments. Assessment Specialist 1 said:

I think it is okay that there must be a unit. Preparing the exams and organizing the thing and you know leading the groups so, it's a good thing. However, in terms of preparing the exams, I think all of the teachers should contribute somehow because they are in the classrooms, and they should contribute to the assessment procedures.

Administering and Scoring Language Assessments. None of the participants talked about a relationship between teaching quality and administering and scoring language assessments.

Evaluating Language Assessments. None of the participants touched upon a relationship between teaching quality and evaluating language assessments.

Academic Leaders' Perception

In this section, first the perceived assessment literacy levels of academic leaders based on the interview data will be presented dimension by dimension. After that, the perceptions of academic leaders on the relationship between assessment literacy and teaching quality will be further explained again based on each dimension.

Perceived Assessment Literacy Level of Academic Leaders

The interviews suggest that academic leaders perceive themselves as both *knowledgeable and skilled* in assessment to a great extent. This high assessment literacy level was asserted by most of the participants, also with reference to the fact that they used to work as assessment specialists at the institution before. They all claimed that they have the necessary knowledge and skills. There was one participant (Academic Leader 1) who considered himself an expert in assessment partially thanks to his higher education in the field. He expressed his assessment literacy level by:

I am specialized in assessment literacy, so I don't want to be very humble about it. I have an MA and also Ph.D. in language assessment. So I can also consider myself as a psychometrician and having specialized in psychometrics, I guess I have enough theoretical knowledge.

While some academic leaders stated that they are better in practice than they are in theory, others claimed the opposite. In this regard, it can be said that the perceptions of academic leaders are similar to the assessment specialists' perceptions on their overall assessment literacy level. For example, Academic Leader 7 said:

I guess I know the general terms in assessment thanks to also the sessions conducted by the testing unit. I see the general structure, how to do it, when to do it, when it is necessary. But of course, I need some practice.

All academic leaders stated that they had taken assessment courses during their education, so they have certain knowledge in the area. On the other hand, especially former assessment specialists including Academic Leader 4 pronounced the lack of theoretical knowledge in assessment although feeling equipped with hands-on practices:

I think I would like to focus more on the theoretical background (of assessment), maybe because here I learned everything for practice. I haven't done my master's in assessment, but in the area, I had classes. I mean, I hadn't done all the courses, so maybe for statistical analysis, I would like to improve.

Firstly, the *knowledge* related dimensions are covered in the following sections.

Knowledge of Theory. Almost half of the seven academic leaders (Academic Leader 3, Academic Leader 5 and Academic Leader 7) shared how assessment and *structure of English* and language learning are related in their perceptions. Therefore, the interview results suggest that some of the academic leaders are knowledgeable about the structure of English as perceived by themselves. Academic Leader 3 acknowledged that language skills should be tested as a whole:

We should know what we aim at as teachers and administrators. We know that our aims are like developing people's language skills as a whole. So our assessment tools should also reflect this objective. Unless they reflect this

objective in the tools, then I don't think we can reach our aims and objectives in the improvement of people's language skills.

Another academic leader, Academic Leader 5, also referred to different skills in a language as he was discussing the assessment practices at the institution as follows:

With multiple choice structure tests, you can test structured knowledge. Then, you can try to test more productive skills with a portfolio. So as a language program, they are focusing on both receptive and productive skills, and we are using a variety of instruments to measure those different skills.

Language Pedagogy. Similar to the English language teachers and assessment specialists, academic leaders also displayed high levels of assessment literacy in the language pedagogy dimensions. All of the academic leaders discussed *washback, feedback and in-class assessment practices* in the interviews. Washback was the most shared theme among the academic leaders, just like among the English teachers. Each academic leader shared their knowledge of the influence of assessment on teaching and learning with different perspectives to be presented more in detail in the next research question. To exemplify, Academic Leader 1, said that the main thing they base their assessment on in curriculum design is the learning objectives. Similarly, Academic Leader 7 believed the main role of assessment is to adapt teaching strategies to empower student learning. One of the academic leaders (Academic Leader 4) shared how she motivates her students through assessment in the following way:

I tell my students about a subject that it's not usually grammatically tested; or be careful, this is in the structure part, especially in this section. So having

that knowledge also I think helps me with the students and they're quite happy about it.

Similarly, Academic Leader 3 recognizes the feedback he gives on the basis of an assessment:

The feedback that we give students, especially for writing is also similar to the feedback we give them for their writing quizzes. That's why the feedback and also the task are similar to the assessment.

Lastly, all of the academic leaders reflected their competence in using different assessment techniques in class such as online assessment. Academic Leader 6 illustrated how she makes use of online assessment tools to promote student learning:

After I teach a topic, I use some extra materials or the online ones. And especially with the online ones, like Quizzes or Quizlet I just download students' responses. And then I check which parts are problematic. So in the next lesson, I can actually focus on those parts or just revise those points.

And after each and every, especially with online months, we look at questions small by one and then analyze them with the extra materials.

The frequency of comments made within this dimension suggests that academic leaders at the institution are quite knowledgeable in language pedagogy dimension of assessment literacy as perceived by themselves.

Principles and Concepts. Concepts of validity and reliability were pronounced by all of the academic leaders at the case institution. For example, Academic Leader 3 thought that he is very knowledgeable in this dimension:

I think I know about the concepts of reliability, what it stands for, and I can give examples of it. and I know about validity and why it's important and

also, use these concepts to evaluate the assessment, tools and items, test items.

Besides, many academic leaders acknowledged strategies to make assessments more valid and reliable. Academic Leader 2 mentioned that specific tasks might be designed to avoid cheating which would negatively affect both validity and reliability. Academic Leader 3 also stated that he is knowledgeable in the concept of inter-rater reliability and gave an example of how it could be achieved:

I think assessing productive skills with another teacher at the same time and then staying there until each person is finished, especially your partner, is good for interrater reliability.

Overall, academic leaders perceived themselves as quite knowledgeable in the fundamental principles of assessment such as validity and reliability.

Impact of Social Values. There was no reference to the impact of social values on assessment practices during the interviews with academic leaders. However, just as with the English teachers and assessment specialists, a few of the academic leaders also mentioned institutional values as an influence on assessment practices at the institution. All of them agreed that as an institutional policy, all assessment-related practices are conducted in a systematic and timely manner.

Local Practices. For this dimension, there were three main categories according to the theoretical framework: teachers' knowledge of accreditation, Turkish assessment traditions and training others to increase assessment literacy. However, none of the academic leaders mentioned *accreditation* during the interviews. However, two academic leaders shared their perceived knowledge of *Turkish assessment traditions* while they evaluated the system at the institution.

Academic Leader 1 stated that many other preparatory schools in Turkey do not prefer subjective assessment as opposed to the case. Academic Leader 3 also shared his knowledge in assessment practices at other preparatory schools in Turkey as follows:

There are proficiency tests in preparatory schools all around Turkey. I don't think they include speaking or writing very much. Because for most institutions the easy thing to mark or score is multiple choice questions or tests. That's why, their proficiency tests do not include productive skills.

Furthermore, a few of the academic leaders shared their knowledge of how they could *train others* in language assessment literacy. Thus, it can be suggested that the academic leaders at the institution are moderately knowledgeable in this dimension based on the interview results.

Personal Beliefs and Attitudes. This dimension focused mainly on awareness of *personal beliefs and attitudes* towards assessment including any *conflicts with other groups* and *the ways to improve one's assessment literacy*. It can be deduced from the interviews that all academic leaders are quite knowledgeable in this dimension because each of them talked about their personal beliefs and attitudes as they answered the interview questions. For instance, Academic Leader 4 mentioned how her teaching philosophy conflicts with some of the assessment practices at the institution by making the following comment:

But here, even when we're talking about speaking, for example, students care more about it until the speaking exam, then, they don't care about it that much. So unfortunately, it also limits the students and the teachers, because you feel that if the students pass, yes, that's an achievement for me, but in my

own philosophy, at least that's not how I feel. Unfortunately, TOEFL being an outdated and limited exam affects all of that.

Similar to the case with assessment specialists and teachers, one academic leader (Academic Leader 6) exemplified how her attitudes toward assessment conflict with the current situation due to the head of the department as follows:

There are so many quizzes or so many concerns. I did not think that testers are in favor of this, but it's because of the head as he actually gives so much emphasis on assessment. This shouldn't be the case. I mean, that person can be the head and also have an expertise in assessment. I mean most of the teachers are complaining about this. Most of the students are complaining about this.

When it comes to the knowledge level of academic leaders in the ways to improve assessment literacy, there were a great number of relevant strategies suggested by each academic leader. Although the results will be further explored in the next research question, a nice summary of suggested ways can be referred to in the following comment made by Academic Leader 1:

So, the first thing they can do is join the free assessment course (provided by the assessment specialists at the institution). Or if they want to specialize further, they can join the testing unit because from time to time, we have got vacant positions. So if they joined the testing unit, then that is also peer learning there. So colleagues are teaching each other. And once you start, there is also a mini orientation program to the testing unit. There is also hands-on training, by the testing specialists unit members here. So that's one way to do it, but if you want to do it academically, then you can do an MA in

educational assessment or in ELT focused on language assessment. You can read the relevant literature as well.

In short, qualitative findings suggest that academic leaders are aware of their personal beliefs and attitudes toward assessment as well as strategies to improve one's assessment literacy to a great extent.

Scores and Decision Making. A few of the academic leaders referred to their decision-making skills. They mostly touched upon their decision-making skills at institutional levels. They stated that assessment helps them make more informed decisions, although some felt that they could manage the process more effectively.

To illustrate, Academic Leader 5 said:

Sometimes we look at the scores to make some critical and related decisions, but we are not doing this on a regular basis. And I think I should also mention that we're not really evaluating our more productive measures, like the portfolio tests for example, we are only focusing on our numeric evaluation and that's what we're only using.

The following sections will present the findings for the *skills* related dimensions.

Constructing Language Assessments. The skills of academic leaders in terms of constructing language assessments were frequently discussed during the interviews. Except for one academic leader, all six academic leaders stated that they can create quality test items for language tests. Due to the fact that four academic leaders (Academic Leader 1, Academic Leader 2, Academic Leader 3, and Academic Leader 4) served as former assessment specialists at the institution, they often reminded their experience as a significant addition to their skills in creating tests.

One academic leader, Academic Leader 2 addressed his skills in multiple-choice type of questions as follows:

For listening to long conversations or talk in the classroom, similar to reading, sometimes the comprehension questions may not be enough to understand. So, before the students listen to the audio, I first listen to it. And then I prepare some questions like multiple choice questions, and then, I give them to students.

Academic Leader 4 also felt quite competent in writing test items as well as the following procedures such as the administration of the test. She said:

So, for generating or writing items, I feel pretty confident. In general, I think in terms of assessment, I know how to like, create most of the items then how to implement them, like how to give and get feedback, like how to make the exam ready and then how to conduct it.

It can be deduced from the qualitative findings that almost all academic leaders, especially the ones with experience as assessment specialists, perceived themselves as very skilled in constructing language test items.

Administering and Scoring Language Assessments. Not many of the academic leaders talked about their skills in scoring language assessments. However, two academic leaders (Academic Leader 7 and Academic Leader 3) said that they are good at scoring writing and speaking assessments. Both of them referred to inter-rater reliability and standardization meetings in the context of grading productive skills.

Evaluating Language Assessments. The qualitative data results suggest that almost half of the academic leaders perceive themselves as very skilled in using statistics to evaluate language assessments. Especially the former assessment

specialists as well as other decision makers demonstrated their skills with relevant examples from the past. Academic Leader 2 and Academic Leader 4 stated that as they worked as assessment specialists before, they improved in item analysis a lot. Furthermore, Academic Leader 5 explained how he can run statistical analyses of test results for evaluation purposes:

I looked at the students in B-level and I compared their test results based on whether they were coming from A level or AR, because we have the assumption that A and AF levels are equal in terms of proficiency. But again, the results show that all A-level students were superior in all tests, compared to AF level students. We can use assessment and the scores we have from there, to understand if our assumptions are accurate. And when you're working in the language field, you may have a lot of assumptions because people always say that there are better ways to do things, but actually doing these analyses for our own micro-environment gives us the possibility to test these assumptions.

Academic Leaders' Perception of the Relationship Between Assessment Literacy and Teaching Quality

The perceptions of academic leaders on the relationship between their assessment literacy and teaching quality showed certain similarities with the perceptions of assessment specialists and teachers. Academic leaders at the institution also frequently discussed *language pedagogy, personal beliefs and attitudes, and principles and concepts* while discussing the relationship between assessment literacy and teaching quality. In addition, perceptions of academic leaders were similar to the perceptions of assessment specialists in that both groups mentioned their skills in *constructing language assessments* as an influential skill in

teaching. However, academic leaders at the institution touched upon a higher number of dimensions of assessment literacy than did the other stakeholders.

To report the findings in more detail, firstly, the *knowledge* related dimensions are covered in the following sections.

Knowledge of Theory. About half of the academic leaders believed that knowledge of theory is a factor to improve teaching quality through assessment literacy. Academic Leader 3, Academic Leader 5 and Academic Leader 7 asserted that assessment should cover all skills in a language for quality assessment. They all referred to different tools in the assessment design of the institution such as quizzes, midterms or portfolios, and how they aim at measuring different aspects of language. In addition, Academic Leader 5 emphasized understanding how students learn is crucial for assessment as learning:

I think it all starts with our understanding of how students learn, how they can create new knowledge, because we're teaching something, this doesn't mean that they are learning it.

In a similar way, Academic Leader 3 referred to students' language skills as well as assessment of language skills as a whole:

We should know what we aim at as teachers and administrators. We know that our aims are like developing people's language skills as a whole. So our assessment tools should also reflect this objective. Unless they reflect this objective in the tools, then I don't think we can reach our aims and objectives in the improvement of people's language skills.

Overall academic leaders believed that the structure of English and how students learn should be taken into consideration when designing assessment tools appropriate for a language curriculum.

Language Pedagogy. In parallel with the results of English language teachers and assessment specialists, academic leaders also talked about their perceptions regarding the relationship between language pedagogy and teaching quality a great number of times. The most frequently appearing categories that emerged during the interviews with academic leaders were washback, student achievement, in-class assessment practices and feedback.

While some of the academic leaders regarded *washback* as a negative effect on their teaching, others believed it helps them shape their classes better. Three of the academic leaders (Academic Leader 1, Academic Leader 2 and Academic Leader 7) stated that assessment results help them reflect on their teaching and adapt their strategies if needed. Since they have additional roles in curriculum design as well, they often referred to how assessment should be used to modify curriculum. For example, Academic Leader 2 said:

If some students, maybe all students consistently get a low grade in some topics, maybe the curriculum unit focuses on this part more in the next semester or next years.

However, a higher number of academic leaders stated that the type of the questions in the exams, too much emphasis on assessments and the proficiency exam TOEFL ITP affected their teaching quality negatively. In this respect, their perceptions were very similar to the English language teachers and assessment specialists. In other words, the interview results suggest that most stakeholders believe that there is a negative washback effect observed at the institution, mainly due to the proficiency exam. One academic leader (Academic Leader 4) explained how the scope of the proficiency exam hinders quality teaching and learning in the following way:

So we have many purposes of assessment but the top priority, the top, number one reason should be actually about the students, but that's not how I feel.

Because of course you need to have an exit exam TOEFL. But still if in the exit exam, if you have equal grading to maybe all the components, like from speaking to writing, then it could change.

One other academic leader (Academic Leader 3), also a former assessment specialist, commented on how productive skills are ignored both in assessment design and in class as follows:

I don't know how to explain but our assessment tools are just tests. Then, we cannot reach these objectives. I don't believe that in my class. It is reflected when you try to do something which requires people to speak or produce or discuss or chat, something like that in language, but people are not very interested in that kind of activity after a certain time and in the institution, this is also reflected when people don't want productive skills.

In a similar manner, some academic leaders discussed the negative impact of some assessments on their teaching. One academic leader (Academic Leader 6) said that many teachers have to teach only certain parts of the book as they are the ones which will be tested. The loaded curriculum, which prevents teachers from taking informed actions based on students' assessment results, was touched upon by other academic leaders as well. For instance, Academic Leader 5 shared his perceptions on how the case institution fails to use assessment results to inform teaching and learning effectively:

We share the results with students. We think that it reflects their progress, but I think we are not using our evaluation to reshape our teaching. No matter what we have on those scores, no matter what is indicated by those scores,

our program tends to stay the same, not very much important by the evaluation practices.

Furthermore, most academic leaders mentioned that English language professionals should refer to assessment results to better understand whether the learning objectives are being achieved by the students. They believed the role of assessment is to provide evidence for student achievement both at personal and institutional levels. One academic leader, Academic Leader 1, believed this is the most essential part of the curriculum in the following comment he made:

Whether the students approach or achieve learning is the most critical thing in the program, and the institution needs to be informed about the learning outcomes, how well the students are doing in relation to the learning outcomes. So, assessments' basic role is how much the students are achieving the learning outcomes.

Academic leaders also emphasized the importance of *feedback* based on assessments to maximize learning and teaching. Although two of them were concerned about the workload and time-constraint issues, they all believed feedback is a crucial part of assessment. Many of them also said that they use their in-class assessment tools to provide feedback for their students' learning. Just like the English language teachers, academic leaders also praised the way they provide feedback for the exams very soon after the administration. Academic Leader 1 gives an example of how the institution is aiming at this as follows:

When we have a formal assessment situation or when we are doing a portfolio assessment, feedback is a critical point there. So, if you have an assessment session on Friday, we provide students with feedback the

following week, and this means that they are also informed how much they have achieved the learning outcomes.

Last but not the least, the use of *in-class assessment practices* to empower learning and teaching was a common concept that emerged during the interviews with academic leaders. The majority of the participants said that they utilize some types of in-class assessments to empower teaching and learning. Similar to the results across all stakeholders, peer-assessment, self-assessment and online assessment were among the most preferred in-class assessments by the academic leaders. To illustrate, Academic Leader 5 said:

I use self-assessment procedures, and peer assessment procedures as much. For example, the self-assessment usually happens in group work while students need to present the information or the content they created with their peers.

Another academic leader (Academic Leader 6) made a nice summary of how academic leaders use online assessment tools to inform teaching and learning in the following way:

After I teach a topic, I use some extra materials or the online ones. And especially with the online ones, like quizzes or quizlet, I just download students' responses. And then I check which parts are problematic. So, in the next lesson, I can actually focus on those parts or just revise those points. And after that, especially with online ones, we look at questions one by one and then analyze them with the extra materials.

However, some continued to discuss the negative washback effect of exams including the proficiency exam on their choice of in-class assessment practices. They said that over years they developed a tendency to focus too much on exercises that

are similar to the exam questions. Some of them said that the Curriculum and Material Development Unit prepares and shares materials which are often TOEFL type questions. Similarly, one academic leader (Academic Leader 3) provided a reflection of his in-class practices affected by the assessments at the institution in the following way:

Even though I don't do something specifically related to assessment, the materials I give them turn out to be related to the assessment. Before the quizzes, you know, we are supposed to give quiz revision and that's also related to the assessment tool. And also before the speaking exams, we do exercises with students for a week and that's also specifically related to the speaking exam. And also the writing tasks that students do weekly. They're also similar to a writing assessment test.

Principles and Concepts. Most of the academic leaders touched upon the validity and reliability of the assessment practices at the institution in the interviews. Overall, everyone agreed that well-designed and fair assessment tools benefit teaching and learning. One academic leader (Academic Leader 5) stated that valid and reliable measures inform teaching goals to a great extent in the following comment he made:

I think assessment holds a very important point in making assumptions about their knowledge by testing the students. We can only measure their performance levels, but those levels, if we use reliable and valid assessment measures, can give us a good idea about the knowledge levels as well.

While describing the strengths of the assessment design at the institution, they said there should be variety in the designs of assessments for different purposes

such as diagnostic assessment, formative assessment, and proficiency testing. One academic leader (Academic Leader 1) made the following comment:

We have lots of different kinds of assessments to make our assessment more reliable and valid. We are not just basing our decisions about students' learning on a single assessment; there's a variety.

However, there were some concerns raised by the academic leaders about the validity and reliability of the portfolio tasks used at the institution. In line with what some English language teachers and assessment specialists also noted down, many academic leaders called for improvement in the design of portfolios. They believed alternative assessment needs to encourage students to involve in teaching and learning processes more. One academic leader discussed how the current situation conflicts with the intended purpose of portfolios as follows:

Most schools in Europe or maybe language courses, use portfolios much more effectively. They (here) give, I think maybe one or two tasks to students in just one semester and they usually copy themselves. So maybe it's not working. The portrait is not working here at our university, but I think it can be more useful.

Another academic leader (Academic Leader 4) also questioned the quality of alternative assessment practices of the institution when she made the following comment:

As for the portfolio, I don't think they're working in this format or in an online format for online homework and portfolio. To be honest to me, they actually seem like we're trying to give students some grades because yes, that's not how we design them. That's not our intention, but the result is actually not representing what we need to do.

Impact of Social Values. Only one academic leader (Academic Leader 3) shared his perceptions related to Turkish assessment systems with specific reference to the design of language proficiency tests in the following manner:

There are proficiency tests of preparatory schools all around Turkey. I don't think they include speaking or writing very much. Because for most institutions the easy thing to mark or score is multiple choice, questions or tests. That's why, their proficiency tests do not include productive skills. And as a result, all students of English, like throughout the country, do not pay specific attention to productive skills.

Local Practices. Unlike the results with the other stakeholders, the most common category discussed for this dimension was training others to improve assessment literacy. Almost half of the academic leaders believed that institutions should be able to provide assessment related training. They believed equipping other teachers with at least the fundamentals of assessment will also encourage them to be more involved in assessment. To this end, assessment trainings, orientation programs, workshops and focus groups were suggested by the academic leaders. One academic leader (Academic Leader 5) who recalled the Peer-Learning training offered by the assessment specialists at the institution summarized his perceptions in the following ways:

They (language professionals) take a regular part in professional development activities, but assessment is more like only a mark. There are some people doing it. They're doing it beautifully. We don't have any responsibilities apart from implementing their instruments and grading. So they don't feel the need to learn more.

Besides, more than half of the academic leaders stated institutions must encourage teachers to improve assessment literacy by providing them with opportunities. One academic leader stated that although experience as an assessment specialist is good, there could be innovative ideas suggested by other teachers out there.

I think it's good that a group of people are specifically charged with that (assessment). But at the same time, I believe that people should also be involved in the discussions in the unit, maybe about what kind of new assessment tools could be used. I mean, getting ideas from people also makes them feel involved in the process. It makes them feel more valued.

Academic Leader 6 made the following comment when she mentioned the Peer-Learning sessions conducted by the assessment specialists:

I think the peer learning part was really effective, but it should be compulsory for the novice ones. How to write reading questions, writing questions, and how they should assess speaking, writing, or in class activities or how they can use the results. Because they do not know, unfortunately, even some experienced ones they'd not care. I think this is the point. I mean, I think there should be more training.

Another academic leader (Academic Leader 2) also explained how training teachers to improve assessment literacy contributes to teaching quality overall as follows:

The training that the testing (unit) delivered last year, peer learning was really, really good because those who are interested in testing and training now have some background information, at least about testing. Each skill was how to write items and also they learned how to do item analysis and et

cetera. So now they have the background information. And maybe those who are not interested in testing may have become interested in testing after the training. I think it's a good way to encourage teachers at the department, but not involve them in the item writing.

Personal Beliefs and Attitudes. The findings from the interviews suggest that academic leaders have some personal beliefs and attitudes which conflict with the real situation at the institution in their perceptions. First of all, they all acknowledged that the system puts a lot of emphasis on assessment practices, which creates stress and decreases teaching and learning quality. In line with what English language teachers and assessment specialists, most academic leaders also stated that there is too much focus on assessment practices due to several reasons one of which is the head of the department being an assessment expert himself. They stated that the assessment design is highly structured, which kills the creativity of the teachers and students. Academic Leader 3 exemplifies how students reflect this too much focus on assessment as follows:

If you teach for the assessment in some way, students see you as a better teacher, because you cater for their needs. But if you don't do things that are not specifically related to the assessment, then you're wasting time.

Other academic leaders also shared their personal beliefs about some assessment practices at the institution which conflict with their teaching philosophies including their teaching strategies. They frequently mentioned the proficiency test which excludes productive skills and how it leads teachers and students to focus on only receptive skills and their assessments. One of them (Academic Leader 6) said that most of the materials provided by the Curriculum Development Unit are in the form of the proficiency test. In the end, many teachers find themselves teaching only

to prepare students for exams and TOEFL ITP. Academic Leader 4 also said this affects student motivation in the following comment she made:

But here, even when we're talking about speaking, for example, students care more about it until the speaking exam, then, they don't care about it that much. So unfortunately, it also limits the students and the teachers. Because you feel that if the students pass, yes, that's an achievement for me, but in my own philosophy, at least that's not how I feel. Unfortunately, TOEFL being an outdated and limited exam affects all of that.

The perceptions of academic leaders in terms of *training ways to improve assessment literacy* were similar to the ones offered by assessment specialists to a great extent. Academic leaders frequently referred to hands-on experience as an effective way to improve assessment literacy. In this respect, many academic leaders mentioned the training sessions offered by the assessment specialists. For instance, Academic Leader 7 made the following suggestion:

Hands-on practice could be a way, or I can just go to some articles and have a look at them. Other than this, I attended the workshops, peer learning can be a good way, or maybe I can just write some items and ask for a designated member friend to check and to give some feedback on them.

Moreover, they said online platforms such as Coursera or YouTube are good ways one can improve their assessment literacy. One academic leader (Academic Leader 2) makes a well-rounded summary of the strategies offered by most academic leaders as follows:

I took some courses on assessment in my master's degree, but of course they are limited. So, I tried to learn some theoretical knowledge about it. I took, for example, online lessons. There are lots of alternatives like Coursera on

other platforms. So, I think online education can be a way and also the different types of education online, like item writing. If you are interested in item writing, you can take some lessons on it. Or if you're just curious about the analysis part, for example, you can just watch some videos on how to conduct analysis on SPSS.

Scores and Decision Making. There were more perceptions of scores and decision-making category as perceived by the academic leaders than by the other English language professionals. The reason could be because they are the ones who define cut-scores and create the academic calendar as well as the assessment schedule. All the academic leaders agreed that assessment results contribute to the quality of decisions regarding the curriculum overall. They said that as academic leaders, they check the assessment results on a regular basis, although this is more difficult with the productive skills. One academic leader explains this with specific reference to the dynamism of the learning and teaching systems:

I think measurement and evaluation together help us make more informed decisions. And we are a very dynamic system. The student profile is always changing. The instructor profile is always changing. We cannot continue to have the same program and they have the same testing style.

The following sections will present the findings for the *skills* related dimensions in assessment literacy.

Constructing Language Assessments. Qualitative findings of this study suggest that most academic leaders perceive teachers' assessment skills in item writing as a contributing instrument to their teaching quality in various ways. All of them believed that a teacher should be competent in writing test items especially with frequent references to course material adaptation and student motivation. As the

majority of the academic leaders were former assessment specialists, they perceived themselves as capable of writing quality test items. Therefore, they could easily adapt the coursebook material thanks to their skills. One academic leader (Academic Leader 6) even said that she has her students practice writing questions for certain content of the book such as reading. One academic leader (Academic Leader 2) explained how he applies his technical skills to serve students better as follows:

For example, we're doing some reading in our classes and sometimes the comprehension questions in the book are not appropriate or to the point. So in some cases I myself prepare some extra questions, comprehension questions, and here, for example, knowing how to write items is quite useful because you can prepare really good items, which serve the students.

Administering and Scoring Language Assessments. Perceptions of academic leaders on this dimension mostly gathered around scoring productive assessments. They said that scoring productive skills such as writing and speaking is an important concept related to assessment, and teachers should be trained on this. Most academic leaders pronounced the effectiveness of standardization meetings managed by assessment specialists for writing exams. To illustrate, Academic Leader 7 made the following comment:

And I can maybe refer to the marking sessions or feedback sessions, I guess it is as objective as possible because it's not only one person's personal judgment. Although there are some criteria, if you read many papers sometimes, there may be some things that you cannot see. So having another person, another colleague to go over the same paper, I guess it's more reliable and also we don't have any place for bias.

Besides, almost all of them agreed that formal assessment practices should be administered by a separate group of people due to the need for specific expertise and also the secrecy issue. In line with the findings that concluded there is an emphasis put on assessment at the case institution, the academic leaders also believed that lack of expertise would negatively affect the validity, reliability and fairness of assessments. To illustrate, Academic Leader 7 said:

I think having a measurement unit is good because it's an important area of teaching. I'm sure that all teachers can add a lot to the units, and do workloads there, but there should be specific things that need to be paid attention to. A mistake that we make in our testing material means a lot. So, I guess having one is safer.

Evaluating Language Assessments. Although the results indicated that academic leaders are very skilled in statistics to evaluate language assessments, not many referred to this as a factor to support teaching and learning. Only one academic leader talked about conducting statistical analyses across different levels in an attempt to compare their proficiency levels. Therefore, it can be said that academic leaders at the institution do not establish a meaningful relationship between teaching quality and technical skills regarding statistics.

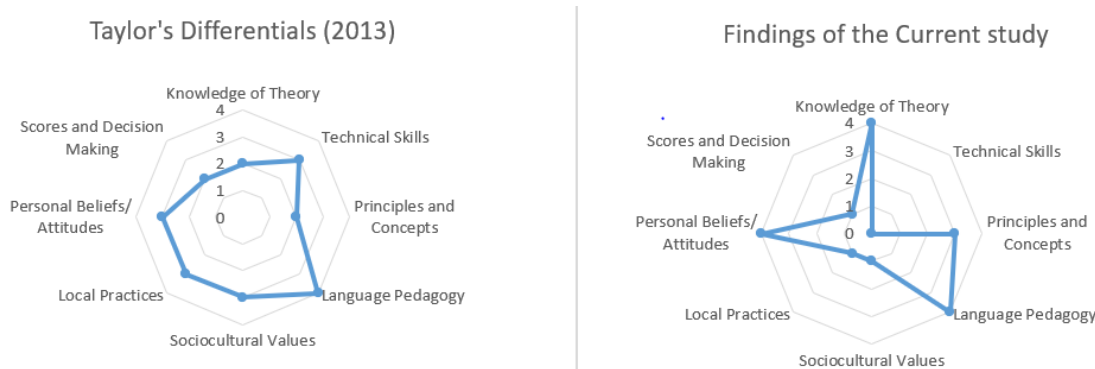
Summary of the Findings

In the light of the survey and interview results, this study aimed to explore assessment literacy perceptions of English language professionals at two levels: English language teachers' assessment literacy levels and the relationship between assessment literacy and teaching quality.

Figure 5 below shows the comparison of the results for teachers' assessment literacy level with the teacher profile in Taylor's framework (2013).

Figure 5

Comparison of the Results with the Differentials in Taylor's Framework (2013)



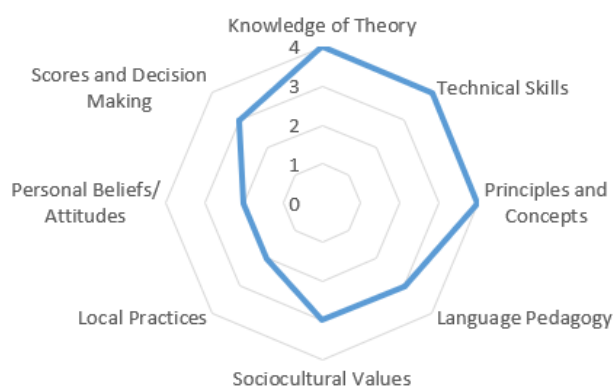
As can be seen in Figure 5, Taylor suggested that *language pedagogy* is the assessment literacy dimension that teachers need the most. The results of the current study also were in line with the framework as teachers' assessment literacy level in language pedagogy was one of the highest. In addition, it was found out that the teachers also perceived high levels of assessment literacy in *knowledge of theory* and *personal beliefs and attitudes* dimensions. One major difference between the teacher profile in the framework and the current study was the results for *technical skills*. Although Taylor's framework claims that teachers require technical skills to a great extent, the results of the current study concluded that teachers do not have assessment literacy in terms of technical skills.

As the current study did not collect quantitative data for the assessment specialists' and academic leaders' profile, the results cannot be represented in a similar fashion in comparison to Taylor's framework as it was done above for the teachers' profile. However, the qualitative results obtained via the interviews with the assessment specialists and the academic leaders will be presented below with reference to the relevant part of the Taylor's framework.

Figure 6 below shows the profile for assessment specialists in Taylor's framework (2013).

Figure 6

Assessment Specialists' Profile in Taylor's Framework (2013)

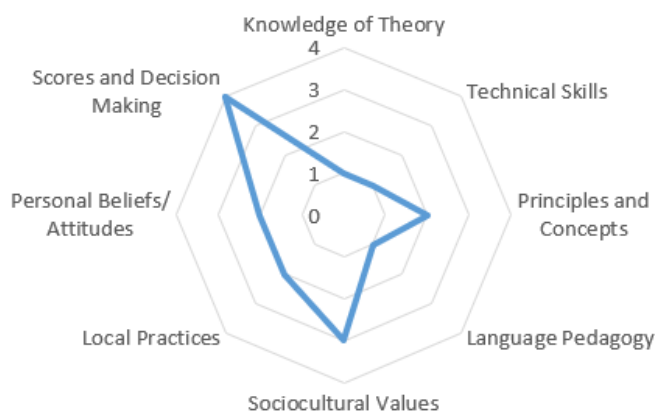


As can be seen in Figure 6, Taylor suggested that *knowledge of theory*, *technical skills* and *principles and concepts* are the assessment literacy dimensions that assessment specialists need the most. On the other hand, *local practices* and *personal beliefs and attitudes* are the ones that assessment specialists need the least. In parallel with the framework, the qualitative results of the current study inferred that the assessment specialists at the institution are knowledgeable in principles and concepts and skilled in the technical skills such as constructing language assessments. One major difference between the framework and the results of the current study was in *personal beliefs and attitudes*. Although Taylor suggested this dimension as one of the skills that assessment specialists need the least, the results of the current study indicated that the assessment specialists are quite knowledgeable in this dimension. In addition, although *knowledge of theory* is one of the most required dimensions according to the framework, assessment specialists who took part in this study did not touch upon this dimension much.

Figure 7 below shows the profile for academic leaders in Taylor's framework (2013).

Figure 7

Academic Leaders' Profile in Taylor's Framework (2013)



As can be seen in the figure, Taylor suggested that *scores and decision making* is the dimension that academic leaders need the most. *Technical skills and language pedagogy* are two dimensions that academic leaders need the least.

However, the current study concluded that the academic leaders at the case institution perceived themselves as quite knowledgeable in all of these dimensions. One reason for this difference could be explained by some of the academic leaders' previous experience as assessment specialists and also their additional responsibilities as teachers. Knowledge of theory, which is another dimension that academic leaders need the least according to the framework, was also found to be low among the academic leaders who participated in the current study.

Table 13 summarizes the perceived assessment literacy levels of three groups of participants. Dimensions with higher levels and dimensions with lower levels or not mentioned dimensions are listed.

Table 13*Summary of Perceived Assessment Literacy Levels Based on Dimensions*

Groups	Data Source	Higher Dimensions	Lower or Not Mentioned Dimensions
Teachers	Survey & Interview	Language pedagogy	Evaluating language assessments
		Knowledge of theory	Scores and decision making
		Personal beliefs and attitudes	Local practices
		Constructing language assessments	Impact of social values
		Principles and concepts	Administering language assessments
Assessment specialists	Interview	Personal beliefs and attitudes	Impact of social values
		Constructing language assessments	Local practices
		Language pedagogy	Knowledge of theory
		Principles and concepts	Evaluating language assessments
		Scores and decision making	Administering language assessments
Academic leaders	Interview	Language pedagogy	Local practices
		Principles and concepts	Impact of social values
		Personal beliefs and attitudes	Evaluating language assessments
		Constructing language assessments	Knowledge of theory
		Scores and decision making	
		Administering language assessments	

Foremost, all stakeholders who participated in this study perceived themselves as quite knowledgeable in *language pedagogy* and *personal beliefs and attitudes*. On the other hand, they are slightly skilled in *evaluating language*

assessments as perceived by all stakeholders. Therefore, the results of the study suggest that English language professionals are *more knowledgeable than skilled* in the assessment literacy dimensions as perceived by themselves. In addition to the survey results, many teachers without additional responsibilities also affirmed in the interviews that they had difficulty in completing the survey as they lacked awareness and/or practice in many of the items. They stated that they lack hands-on experience and/or skills, especially when it comes to preparing exams and conducting statistical analyses on assessments. Although most of them had assessment courses during their studies, they did not have a chance to gain first-hand experience, partially because there is a specific unit managing the assessment processes at the case institution. In this respect, *using statistics* was found to be the category that many stakeholders perceive as their lowest assessment literacy level. However, both assessment specialists and academic leaders perceived themselves better in hands-on skills such as item writing or getting an exam ready to deliver. Only the participants with related experience, assessment specialists and academic leaders, acknowledged their skills in scores and decision-making. In a similar respect, academic leaders and assessment specialists said during the interviews that they perceive themselves as more assessment literate than most other teachers at the institution. They frequently referred to their additional responsibilities when they discussed their assessment literacy levels. Overall, the findings of this study suggest that academic leaders and assessment specialists are more knowledgeable and skilled in assessment literacy than the teachers without additional responsibilities.

Table 14 below presents the frequency of references to assessment literacy dimensions in relation to teaching quality. Frequently referred dimension and less or not referred dimensions are listed.

Table 14*References to the Assessment Literacy Dimensions in relation to Teaching Quality*

Group	Frequently Referred Dimensions	Less or Not Referred Dimensions
Teachers	Language pedagogy Personal beliefs and attitudes Principles and concepts	Administering language assessments Evaluating language assessments Scores and decision making Impact of social values Knowledge of theory Constructing language assessments Local practices
Assessment specialists	Language pedagogy Personal beliefs and attitudes Constructing language assessments Scores and decision making Principles and concepts	Knowledge of theory Administering language assessments Impact of social values Local practices Evaluating language assessments
Academic leaders	Language pedagogy Personal beliefs and attitudes Constructing language assessments Principles and concepts Knowledge of theory Scores and decision making Local practices	Impact of social values Administering language assessments Evaluating language assessments

As can be seen in Table 14, perceptions of stakeholders on the relationship between assessment literacy and teaching quality were both similar and different to a certain extent. The common dimensions discussed by all stakeholders were *language pedagogy*, *personal beliefs and attitudes*, and *principles and concepts*. In other words, English language professionals all agreed that these dimensions of assessment literacy have a strong relationship with teaching quality. On the other hand, there

were some variations among the participants' perceptions on some dimensions. For instance, although none of the teachers discussed *decision-making* skills with reference to assessments during the interviews, both academic leaders and assessment specialists touched upon its essence for the relevant decisions regarding curriculum design. Similarly, *constructing language assessments* dimension was discussed much more by assessment specialists and academic leaders than by the teachers. This is an expected and predictable perception difference considering the circumstances of the case institution where specific people are in charge of curriculum and assessment designs and the relevant decision-making processes.

Table 15 presents the most pronounced concepts of several assessment literacy dimensions as perceived by English language professionals.

Table 15

Most Frequently Covered Concepts of Assessment Literacy Dimensions

Dimensions	Concepts	Perceptions
Language pedagogy	Washback	All stakeholders
	In-class assessments	Teachers and academic leaders
	Feedback	Teachers and academic leaders
Constructing language assessments	Item writing	Assessment specialists and academic leaders
Personal beliefs and attitudes	Proficiency exam	All stakeholders
	Exam-oriented culture	All stakeholders
	Training	All stakeholders
Principles and concepts	Validity	All stakeholders
	Reliability	All stakeholders

Overall, the participants agreed that English language professionals should improve their assessment literacy so as to promote teaching and learning. Most of them said assessment practices should inform teaching goals, and in this respect,

teachers should have an overall understanding of how assessments can influence instruction. In this context, some English language professionals talked about *washback* as a positive influence on their teaching. Moreover, they emphasized the importance of using *in-class assessment tools* to support teaching and learning. Most participants said that they use peer-assessment, self-assessment, and online-assessment in their classes to see their students' levels, give feedback and modify instruction. In addition to using ready-made assessment tools, some also mentioned the benefits of constructing language items to support their instruction when needed. Therefore, *item writing* was an important skill that increases English language professionals' teaching quality through assessments as perceived by all stakeholders. On the other hand, some stakeholders felt that there is too much importance put on assessment, which hinders teaching and learning. They also referred to the *proficiency exam*, which is a high-stake language test, TOEFL ITP, as a factor that causes negative washback. In addition, the results indicate that English language professionals question the *validity* and *reliability* of the assessments because of this negative effect on both students and teachers. They often stated that the absence of productive skills in TOEFL ITP negatively affects the design and implementation of some assessments including the portfolio. As a consequence, teachers, assessment specialists and academic leaders all suggested various *training* methods to improve the assessment literacy knowledge and skills of English language professionals. The suggestions include but are not limited to hands-on workshops, reading relevant literature, and pursuing a degree in assessment.

CHAPTER 5: DISCUSSION

Introduction

In this chapter, firstly, an overview of the study is given. After that, discussion of major findings related to the assessment literacy perceptions of English language professionals are presented with reference to the relevant literature. Implications for practice as well as for further research, and the limitations of the study are also provided.

Overview of the Study

One of the purposes of this study was to explore the assessment literacy levels of tertiary level English language teachers perceived by themselves, by assessment specialists, and by academic leaders. The other purpose was to explore how tertiary level English language professionals perceive the relationship between assessment literacy and teaching quality. In the first stage, a Likert-type survey was distributed to the participants and the results displayed the perceived assessment literacy levels of English language teachers. Following the survey, volunteered teachers, assessment specialists and academic leaders were interviewed to collect data on their perceptions of assessment literacy levels and its relationship to teaching quality. Both data sets were analyzed and interpreted to discuss conclusions about assessment literacy perceptions of tertiary level English language professionals in this case study.

Discussion of Major Findings

English Language Teachers' Assessment Literacy Level

The quantitative results of this study indicated that English language teachers are perceived as moderately knowledgeable and slightly skilled in the assessment

literacy dimensions of Taylor's framework (2013). The perceived levels were much lower in dimensions of assessment literacy skills than those of assessment literacy knowledge. Although it was noted that many English language teachers have taken courses related to assessment during their studies, they did not feel self-assured in assessment wise decisions and practices due to lack of experience and expertise. Several other researchers studying English language teachers' assessment literacy levels through various tools also found a low level of assessment literacy among English language teachers (Fulcher, 2012; Kiomrs et al., 2011; Talib et al., 2013). For instance, Muhammad and Bardakçı (2019) asserted that Iraqi English language teachers' assessment literacy level was very low, especially in the use of assessment for learning. Similarly, Talib et al. (2013) found out that English teachers in Malaysia were neither knowledgeable nor skilled in fundamental concepts of language assessment, and thus they had low levels of assessment literacy. Results were similar in the Turkish context too. In one study, Hatipoğlu (2015) focused on participants' literacy levels in classroom-focused language assessment as well as concepts of assessment literacy and found that English language teachers in Turkey were not competent enough in language testing and assessment (as cited in Şişman & Büyükkarcı, 2019). Similar to the participants of the current research, English language teachers demonstrated inadequate knowledge of assessment and a lack of assessment literacy skills. Tsagari and Vogt (2017) also explored the perceived assessment literacy levels of European teachers of the English language and noted similar results. Their study also suggested that the perceived assessment literacy levels of the teachers were low, and not adequate to perform assessment related practices in an effective way.

Language pedagogy dimension, which included several items related to teachers' in-class assessment practices and washback effect in Kremmel and Harding's tool (2020) used in the study, was the highest level perceived by teachers. This conclusion is meaningful within the theoretical framework (Taylor, 2013) as the framework suggested that language pedagogy is the most essential dimension for teachers. Therefore, English language professional need to improve their assessment literacy for valid and reliable assessment for learning (Atmojo, 2021; Buckley-Walker & Limpscombe, 2021; Davidson, 2018; Earl, 2013; Kremmel & Harding, 2020). On the other hand, it was concluded that teachers without any additional responsibilities had lower level of assessment literacy compared to assessment specialists and academic leaders, especially in terms of the assessment literacy skills as perceived by all stakeholders. This was an expected result as certain processes are managed by certain people who are in charge at the case institution. Thus, the teachers did not have experience in some areas of assessment such as conducting *statistical analyses for assessments*. Some also stated that they do not feel the urge to be competent in certain areas, mostly because they are not the ones responsible for managing those assessment-related processes such as creating rubrics. These findings are in line with those of previous studies. Coombs et al. (2021) found out that teachers who have experience in leadership positions at institutions have a more comprehensive understanding of assessment approaches, and they value assessment and assessment related data as a major contributor to improving teaching and learning more than those in non-leadership positions. Taylor (2009) also emphasized that a balance of theoretical knowledge and practical skills need to be provided for language teachers, and this balance is to be affected by the nature of the organization such as the level, nature and type of involvement of teachers in the assessment

processes. With the increasing involvement and interest of teachers in the assessment processes (Coombe et al., 2020), it can be suggested that English language teachers need further training on assessment literacy to increase their assessment literacy dimension based on their roles and job descriptions. The lack of training, pronounced by the participants of the current study, was also supported by the findings of previous studies. For instance, Xu and Brown (2017) indicated that the EFL instructors at Chinese universities had a low level of language assessment literacy, mainly due to a lack of in-service training. Fulcher (2012) also found out that the level of English teachers' assessment literacy was not high, and that this mostly resulted from a lack of training and practice in the area.

Another commonly covered concept which could be an explanation for the low levels of assessment literacy of English teachers was the proficiency test, TOEFL ITP administered in the case school. The participants believed that having a standardized test affected their assessment literacy level in a negative way. The perceived effect of high-stake exams as hindrance on assessment literacy level had previously been explored by other researchers (Fan & Yang, 2020; Green, 2006; Wall, 2012), and the findings of the current study are in line with this line of research. For instance, Kiomrs et al. (2011) found out that the use of a standardized test was pushback in teachers' assessment literacy in his study where he worked with Iranian teachers of the English language in their studies and revealed low levels of language assessment literacy.

The Relationship Between Assessment Literacy and Teaching Quality

The major conclusions drawn from this study suggest that English language professionals perceived assessment literacy as a major contributor to teaching quality to a great extent. Overall, they believed that a high level of assessment literacy could

contribute to teaching quality through better interpretation of the assessments used in language teaching, motivating student learning and achievement through in-class assessments, constructive feedback based on assessments and so on so forth.

Relevant literature also confirms that teachers with language assessment literacy facilitate their students' assessment related practices more effectively in various ways and empower teaching and learning better (Coombe et al., 2020; Green, 2018; Rea-Dickins, 2006; Stiggins, 2002). The perceptions of teachers, assessment specialists and academic leaders were similar to the extent that relates to teaching practices as all groups of participants are teachers at the core in addition to some other responsibilities. For instance, all English language professionals perceived *language pedagogy* and *personal beliefs and attitudes* dimensions as main reference points in relation to teaching quality. However, some dimensions such as *evaluating language assessments* and *local practices* were rarely recognized by teachers, assessment specialists and academic leaders. Only a few participants who had relevant job descriptions were insightful about these dimensions, probably due to their additional responsibilities at the case institution. Considering the framework of the study (Taylor, 2013), this perception difference is intelligible as different stakeholders may have different levels of demands and focuses based on the LAL requirements based on their professional roles (Kremmel & Harding, 2020).

The current study also concluded that English language professionals established a strong relationship between their assessment literacy perceptions of washback and their teaching quality. Their personal beliefs and attitudes towards assessment practices and specifically washback effect were found to affect their teaching quality across all stakeholders. The most notable reasons behind this perception were having a high-stake exam (TOEFL ITP) as the proficiency exam and

ignoring productive skills because of exam-oriented instruction which leads to concerns about the validity and reliability of assessment practices. The participants commonly stated the proficiency exam leads the students to memorize commonly asked items, and prompts “unproductive learning”. These results are parallel with the line of relevant research in literature. Other researchers also suggested a negative effect of high-stake exams on teachers’ assessment literacy. For instance, participants of a study conducted by Tsagari (2011) stated that the use of a high-stake exam launched by Cambridge ESOL (English for Speakers of Other Languages) affected teachers’ assessment literacy and their teaching quality in a negative way. The current study revealed that English language teachers tended to focus too much on multiple-choice type of questions for their classroom assessments since the proficiency exam consists of multiple-choice type of questions (Kiomrs et al., 2011; Llosa, 2011; Yildirim & Orsdemir, 2013). The current situation did not reflect the intended teaching and learning goals in their teaching philosophies. Both the teachers and the students tended to ignore the importance of productive skills such as writing and speaking because they are not included in the proficiency exam, TOEFL ITP. Other studies also yielded similar results regarding the negative effect of traditional assessment and test types on teachers’ teaching quality. To illustrate, Alqahtani (2021) stated that language proficiency tests such as the TOEFL or the IELTS avoid the stakeholders from focusing on different aspects of the language that are not covered in the high-stake tests. Similarly, Tsagari and Vogt (2017) concluded in their study that European teachers of English tend to focus on traditional test types such as multiple-choice questions as they had a similar proficiency test. Therefore, it can be concluded that having a standardized test, a high-stake exam or a traditional proficiency exam with multiple choice type of questions may decrease the

importance given to the assessment of productive skills which are not tested in those tests such as speaking and hinders engagement in those skills in and out of class.

Although the teachers are aware of alternative assessment methods such as portfolios, their low level of assessment literacy as well as the nature of the assessments used at school can cause them not to make use of assessment for teaching and learning methods. As a result, formative assessment and alternative assessment such as portfolios should be better integrated into the curriculum of language schools with an extended purpose to increase teachers' assessment literacy of assessment for learning (Fathi & Rahimi, 2022; Green, 2018; Jia et al., 2006; Nasab, 2015; Stiggins, 2014).

English language professionals recognized the effect of their assessment literacy regarding positive washback on their teaching to the extent that they can modify instruction based on assessment results. Furthermore, they perceived in-class assessment tools as essential to their teaching, and some even applied item-writing skills to exercise in-class assessment and assessment for learning. The use of online assessment was also commonly mentioned as a contributor to their assessment literacy and teaching quality. These results of the current study are in line with the previous studies. Many other researchers in the relevant literature also suggested that the use of formative assessments helps teachers shape their teaching according to the assessment results; and assessment literacy enables them to make use of assessment practices with a learning-oriented approach (Buckley-Walker & Lipscombe, 2021; Coombs et al. 2021; Giraldo, 2019; Hamp-Lyons, 2017).

One last major finding of the current study was that English language professionals called for more training to improve their assessment literacy. Although the needs raised varied across participants, they agreed that training on assessment

should be offered to English language professionals to help them better perform their jobs. The need for training in assessment literacy has been frequently covered by the relevant literature in the past decade. To start with, as the pioneer of the concept of assessment literacy, Stiggins (2014) continuously emphasized the importance of training to improve teachers' assessment literacy and stated that schools and workplaces need to create an environment where teachers have access to training and practice in assessment so as to empower learning and teaching. The participants of the current study also commonly stated that they should be provided with more training on assessment knowledge and skills by the institution. They believed that training on assessment is crucial for their professional development as well as the quality of their teaching. Taylor (2013) also asserted that developing assessment literacy through various types of training resources is considered substantial for the core knowledge and skills that base good quality assessment and teaching widely. With the increased use of assessments for various purposes, and the increased responsibility given to teachers, English language teachers would need more training on assessment to increase their assessment literacy and make their teaching more informed and effective (Bijsterbosch et al., 2019; Fulcher, 2012; Hidri, 2021; Hughes, 2021; Tsagari & Vogt, 2017). Therefore, it can be said that these teachers should be provided with more guidance and training to expand this positive washback in certain areas of their job.

The assessment literacy needs shared by the assessment specialists and academic leaders differed from those shared by the teachers due to different profiles. While the teachers considered they should improve in their use of in-class assessment practices to increase student achievement and motivation as major aspects to their teaching quality, assessment specialists and academic leaders further

required several technical skills such as statistical analyses of language assessments as they take active roles in these domains. This seemed to be in parallel with the framework used in the study (Kremmel & Harding, 2020; Taylor, 2013). As a result, it can be concluded that stakeholders should be provided with training on assessment literacy based on the diverse nature and job descriptions of the stakeholders (Stiggins & Duke, 2008; Stiggins, 2014; Şişman & Büyükkarcı, 2019; Tsagari, 2017).

Implications for Practice

Several recommendations are noteworthy in the light of the findings of this study. To start with, all stakeholders at the institution perceived assessment literacy as fundamental to quality teaching. Most of them were aware of how assessment practices influenced teaching and learning to a certain extent. However, a great number of participants also mentioned certain factors such as the TOEFL ITP or exam-oriented culture as a hindrance to assessment as learning. Therefore, it can be suggested that English language professionals are provided with *training on washback* effect to minimize the potential negative effect of such exams on learning and teaching.

Secondly, almost all stakeholders agreed that assessment literacy adds to teaching and learning in different ways. On the other hand, many of them, especially the ones without additional responsibilities stated that they are not very familiar with the concepts of assessment literacy. Due to the high number of participants who also said the survey was hard to complete as they had never thought about those dimensions before, it can be said that assessment literacy is relatively a new concept for the English language teachers at the institution. To this end, more theoretical and practical aspects of assessment could be offered through undergraduate and graduate level assessment courses. Assessment literacy among English language professionals

could be improved by means of relevant courses provided during pre-service and in-service teacher education degree programs. In addition, academic leaders and professional development units could *encourage English language professionals to expand their assessment literacy* by providing them with the opportunities such as access to theoretical knowledge, creating favorable circumstances where teachers can gain and practice and know-how skills, continuous professional development activities, workshops and so on so forth. An effective way for the administrators and professional development units to motivate English language professionals to actively benefit from these opportunities could be micro-credentials and/or digital badges offered to the participants upon training. Presented to English language professionals at the end of assessment related professional development activities, these certifications could affirm recognition and participation. Furthermore, schools can collaborate with associations like DEDAK, which aims to enhance the quality of language education in Turkey through accreditation and evaluation studies, in accordance with this purpose.

Both quantitative and qualitative results also concluded that English language teachers without experience as an assessment specialist lacked technical skills such as constructing language assessments. In the same respect, quite many participants asserted the impact of hands-on practices on improving assessment literacy and assessment as learning as well. Several suggestions from the interviews were reconducting Peer-Learning training, creating a pool for assessment methods for the teachers to contribute to, and lastly involving teachers more in the assessment processes in general. To this end, English language professionals could be offered *hands-on practices* such as training provided by internal and external experts in field,

creating practical solutions based on the circumstantial requirements of the institutions.

One last conclusion drawn was that almost all teachers believed productive skills were not assessed at their best, partly because of their perceived role in the curriculum. Many teachers mentioned their concerns about the validity and reliability of productive skills assessment. Many participants from different stakeholder groups also repeated that the current portfolio design did not serve its purpose, and thus needed revision. As a result, academic leaders, assessment specialists and English language teachers at tertiary level institutions should collaborate should call for *modification of assessment design* in a way that it addresses all skills and aspects of language more forcefully.

Implications for Further Research

As the current study was conceptualized as a case study, the number of participants was limited to the current institution. As an implication for further research, it can be suggested that assessment literacy perceptions of English language professionals at other institutions could be explored for generalizable results. With higher numbers of participants, inferential statistics could also be conducted to see the correlation of age, department or year of experience with the perceptions of assessment literacy. Such a study could also indicate whether the amount and nature of the participants' theoretical and practical input during their pre-service and/or in-service teacher education degree programs have an impact on their assessment literacy.

Although the current study encompassed several dimensions in Taylor's framework (2013), the washback effect was obviously the most recognized concept discussed in the interviews. Thus, further research could specifically explore

perceptions on assessment literacy and positive washback through a more targeted design.

Considering the framework, this study involved teachers, assessment specialists and academic leaders as participants. Further research could involve students as they are core stakeholders in the learning process. Therefore, a study on students' perceptions of assessment literacy could provide valuable insight to the relevant literature.

Finally, as most participants acknowledged the relationship between assessment literacy and teaching, further research on how English language professionals can improve their assessment literacy to allow assessment for/as learning can yield fruitful results for all stakeholders. To illustrate, a theoretical review of *assessment for/as learning* with an aim to explore effective practices to increase assessment literacy would be beneficial.

Limitations of the Study

One limitation of this study was that a limited number of English language professionals working in the case institution were involved in this study. Although all of the academic leaders and assessment specialists participated in the study, only less than half of the teachers were involved. Therefore, the results might not be reflecting the perceptions of all teachers in the case institution. Moreover, due to the low number of the surveys filled in, inferential statistics could not be conducted. Therefore, the results cannot be generalized to other institutions (Bakeman & Robinson, 2005). Finally, within the limitations of the framework, the study did not include the perceptions of students, which could have explored the phenomena more profoundly if added.

REFERENCES

- Ahmed, F., Ali, S., & Shah, R. A. (2019). Exploring variation in summative assessment: Language teachers' knowledge of students' formative assessment and its effect on their summative assessment. *Bulletin of Education and Research*, 41(2), 109-119. <https://files.eric.ed.gov/fulltext/EJ1229441.pdf>
- Aksit T., Mengü H. I. & Turner R. (Eds.). (2018). *Bridging teaching, learning and assessment in the English language classroom*. Cambridge Scholars.
- Alqahtani, F. (2021). The impact of language testing washback in promoting teaching and learning processes: A theoretical review. *English Language Teaching*, 14(7), 21-26. <https://files.eric.ed.gov/fulltext/EJ1309575.pdf>
- Atmojo, A. E. P. (2021). Facilitating positive washback through authentic assessment in EFL assessment. *Journal of Teaching and Education*, 3(2), 226-233. <https://doi.org/10.30650/ajte.v3i2.2219>
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford University Press.
- Bakeman, R., & Robinson, B. F. (2005). *Understanding statistics in the behavioral sciences*. Psychology Press.
- Bashir, S., Karim, T., & Akram, N. (2020). Teachers focus to practice formative assessment techniques and Blooms Taxonomy: Assessment of class practices of prospective teachers. *Journal of Arts & Social Sciences (JASS)*, 7(2), 116-125. [https://doi.org/10.46662/jass-vol7-iss2-2020\(116-125\)](https://doi.org/10.46662/jass-vol7-iss2-2020(116-125))
- Bergeron, L. (2020). An investigation into the relationships among middle school teachers' beliefs about collaboration, their perceptions of formative assessment, and selected teacher characteristics. *Current Issues in Education*, 21(3), 1-18. <https://cie.asu.edu/ojs/index.php/cieatasu/article/view/1714>

- Bijsterbosch, E., Béneker, T., Kuiper, W., & van der Schee, J. (2019). Teacher professional growth on assessment literacy: A case study of prevocational geography education in the Netherlands. *The Teacher Educator*, 54(4), 420-445. <https://doi.org/10.1080/08878730.2019.1606373>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31.
<https://link.springer.com/content/pdf/10.1007/s11092-008-9068-5.pdf>
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675. <https://doi.org/10.2307/3587999>
- Buckley-Walker, K., & Lipscombe, K. (2021). Validity and the design of classroom assessment in teacher teams. *The Australian Educational Researcher*, 49(2), 425-444. <https://link.springer.com/content/pdf/10.1007/s13384-021-00437-9.pdf>
- Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., & Walker, K. (2020). Purposive sampling: complex or simple? Research case examples. *Journal of Research in Nursing*, 25(8), 652-661.
<https://doi.org/10.1177/1744987120927206>
- Chapelle, C. A., (2012). Conceptions of validity. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 21-33). Routledge.
- Chirimbu, S. (2013). Using alternative assessment methods in foreign language teaching. Case study: Alternative assessment of business English for university students. *Buletinul Stiintific al Universitatii Politehnica din Timisoara, Seria Limbi Moderne*, 12(1-2), 91-99.

https://sc.upt.ro/images/cwattachments/104_39c8127534353f27dae09c5cbcae_d408.pdf

Collier, J. K., & Huang, B. (2020). Test review: Texas English language proficiency assessment system (TELPAS). *Language Assessment Quarterly*, 17(2), 221-230. <https://doi.org/10.1080/15434303.2019.1691216>

Connors, C. B. (2021). Summative and formative assessments: An educational polarity. *Kappa Delta Pi Record*, 57(2), 70-74. <https://doi.org/10.1080/00228958.2021.1890441>

Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia*, 10(3), 1-16. <https://doi.org/10.1186/s40468-020-00101-6>

Coombs, A., Rickey, N., DeLuca, C., & Liu, S. (2021). Chinese teachers' approaches to classroom assessment. *Educational Research for Policy and Practice*, 21(1), 1-18. <https://doi.org/10.1007/s10671-020-09289-z>

Creswell, J. W. & Creswell, J. D. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research* (3rd ed.). Sage Publications.

Crisp, G. (2011). *Teacher's handbook on e-assessment*. Australian Learning and Teaching Council Limited.

Cumming, A. (2009). Language assessment in education: Tests, curricula, and teaching. *Annual Review of Applied Linguistics*, 29, 90-100. <https://doi.org/10.1017/S0267190509090084>

- Cumming, A. (2013). Assessing integrated writing tasks for academic purposes: Promises and perils. *Language Assessment Quarterly*, 10(1), 1-8.
<https://doi.org/10.1080/15434303.2011.622016>
- Davidson, P. (2018). How can we use assessment to facilitate learning? In T. Akşit, H. I. Mengü & R. Turner (Eds.), *Bridging teaching, learning and assessment in the English language classroom* (pp. 8-16). Cambridge Scholars.
- Djoub, Z. (2017). Assessment literacy: Beyond teacher practice. In R. Al-Mahrooqi, C. Coombe, F. Al-Maamari & V. Thakur (Eds.), *Revisiting EFL assessment: Critical perspectives* (pp. 9-27). Springer.
- Earl, L. M. (2013). Assessment for learning; assessment as learning: Changing practices means changing beliefs. *Assessment and Learning*, 80(2), 63-71. [https://wlts.edb.hkedcity.net/filemanager/file/AandL2chapter/A&L2_\(1\)%20Lorna.pdf](https://wlts.edb.hkedcity.net/filemanager/file/AandL2chapter/A&L2_(1)%20Lorna.pdf)
- Elbow, P., & Belanoff, P. (1986). Portfolios as a substitute for proficiency examinations. *College Composition and Communication*, 37(3), 336-339.
<https://doi.org/10.2307/358050>
- Elder, C. (2017). Language assessment in higher education. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (3rd ed., pp. 271-287). Springer.
- Fan, H., & Yang, M. (2020). The washback effect of English tests on language learning strategies: A comparative case study based on the data analysis of the English and non-English majors in China. *Proceedings of 2020 International Conference on Computer Information and Big Data Applications (CIBDA)* (pp. 41-47). IEEE.
<https://doi.org/10.1109/CIBDA50819.2020.00018>

Farahian, M., Avarzamani, F., & Rajabi, Y. (2021). Reflective thinking in an EFL Writing course: To what level do portfolios improve reflection in writing? *Thinking Skills and Creativity*, 39, 1-12.

<https://doi.org/10.1016/j.tsc.2020.100759>

Fathi, J., & Rahimi, M. (2022). Electronic writing portfolio in a collaborative writing environment: Its impact on EFL students' writing performance. *Computer Assisted Language Learning*. Advance online publication.

<https://doi.org/10.1080/09588221.2022.2097697>

Felder, R. M., & Brent, R. (1999). How to improve teaching quality. *Quality Management Journal*, 6(2), 9-21.

<https://doi.org/10.1080/10686967.1999.11919183>

Fergus, S., Hirani, E., Parkar, N., & Kirton, S. B. (2021). Strategic engagement: Exploring student buy-in across a formative and summative online assessment. *All Ireland Journal of Higher Education*, 13(1), 1-24.

<https://ojs.aishe.org/index.php/aishe-j/article/view/441/889>

Fox, J. (2017). Using portfolios for assessment/Alternative assessment. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (3rd ed., pp. 135-149). Springer.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education*. McGraw Hill.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment* (1st ed.). Routledge.

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

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132.

<https://doi.org/10.1080/15434303.2011.642041>

Gay, L. R., Mills, G. E., & Airasian, P. W. (2009). *Educational research competencies for analysis and applications*. Merrill/Pearson.

Gholson, M. L., & Guzman-Orth, D. (2019). Developing an alternate English language proficiency assessment system: A theory of action. *ETS Research Report Series*, 2019(1), 1-19.

<https://doi.org/10.1002/ets2.12262>

Giraldo, F. (2019). Language assessment practices and beliefs: Implications for language assessment literacy. *How*, 26(1), 35-61.

<https://doi.org/10.19183/how.26.1.481>

Giraldo, F. (2020). A post-positivist and interpretive approach to researching teachers' language assessment literacy. *Profile Issues in Teachers Professional Development*, 22(1), 189-200.

<https://doi.org/10.15446/profile.v22n1.78188>

Gould, S. J. (1981). *The mismeasure of man*. Norton & Company.

Green, A. (2006). Watching for washback: Observing the influence of the international English language testing system academic writing test in the classroom. *Language Assessment Quarterly*, 3(4), 333-368.

<https://doi.org/10.1080/15434300701333152>

Green, A. (2018). Teacher assessment literacy for the classroom. In T. Aksit, H. I. Mengü, & R. Turner (Eds.), *Bridging teaching, learning and assessment in the English language classroom* (pp. 2-8). Cambridge Scholars.

- Green, A. (2021). Exploring language assessment and testing: Language in action. *ELT Journal*, 69(1), 109-110. <https://doi.org/10.1093/elt/ccu065>
- Guo, Q., & Xu, Y. (2021). Formative assessment use in university EFL writing instruction: A survey report from China. *Asia Pacific Journal of Education*, 41(2), 221-237. <https://doi.org/10.1080/02188791.2020.1798737>
- Guzman-Orth, D., Laitusis, C., Thurlow, M., & Christensen, L. (2016). Conceptualizing accessibility for English language proficiency assessments. *ETS Research Report Series*, 2016(1), 1-12. <https://doi.org/10.1002/ets2.12093>
- Hamp-Lyons, L. (2017). Language assessment literacy for language-oriented assessment. *Papers in Language Testing and Assessment*, 6(1), 88-111. <http://hdl.handle.net/10547/622445>
- Hatipoğlu, Ç. (2015). *Diversity in language testing and assessment literacy of language teachers in Turkey* [Paper presentation]. The 3rd ULEAD Congress, International Congress on Applied Linguistics: Current Issues in Applied Linguistics, Çanakkale, Turkey.
- He, W., & Mei, Q. (2020). Research into formative assessment in English writing for English majors. *Proceedings of 2020 International Symposium on Advances in Informatics, Electronics and Education (ISAIEE)* (pp. 50-53). IEEE. <https://doi.org/10.1109/ISAIEE51769.2020.00019>
- Herrera Mosquera, L., & Macías V, D. F. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal*, 17(2), 302-312. <https://doi.org/10.14483/udistrital.jour.calj.2015.2.a09>

- Hidri, S. (2021). Language assessment literacy: Where to go? In S. Hidri (Ed.), *Perspectives on language assessment literacy: Challenges for improved student learning* (1st ed., pp. 3-12). Routledge.
- Holt, D. D. (1992). *Alternative approaches to assessment & evaluation in family English literacy programs*. California Department of Education.
- Huang, H. T. D., & Hung, S. T. A. (2017). EFL Test-takers' feedback on integrated speaking assessment. *TESOL Quarterly*, 51(1), 166-179.
<https://www.jstor.org/stable/44986982>
- Hughes, A. (1989). *Testing for language teachers*. Cambridge University Press.
- Hughes, L. (2021). Washback and the assessment practices of ESL instructors at Japanese university: Language literacy. *Journal of Linguistics, Literature, and Language Teaching*, 5(1), 1-9. <https://doi.org/10.30743/ll.v5i1.3238>
- Jia, Y., Eslami, Z. R., & Burlbaw, L. M. (2006). ESL teachers' perceptions and factors influencing their use of classroom-based reading assessment. *Bilingual Research Journal*, 30(2), 407-430.
<https://doi.org/10.1080/15235882.2006.10162883>
- Jiang, Y. (2020). Teacher classroom questioning practice and assessment literacy: Case studies of four English language teachers in Chinese universities. *Frontiers in Education*, 5(23), 1-17.
<https://doi.org/10.3389/feduc.2020.00023>
- Jones, N. (2012). Reliability and dependability. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 350-362). Routledge.
- Kane, M. (2012). Articulating a validity argument. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 34-47). Routledge.

- Kaplan, B., & Maxwell, J. A. (2005). Qualitative research methods for evaluating computer information systems. In J. G. Anderson & C. E. Aydin (Eds.), *Evaluating the organizational impact of healthcare information systems* (pp. 30-55). Springer.
- Kiomrs, R., Abdolmehdi, R., & Rashidi, N. (2011). On the interaction of test washback and teacher assessment literacy: The case of Iranian EFL secondary school teachers. *English Language Teaching*, 4(1), 156-161.
<https://files.eric.ed.gov/fulltext/EJ1080409.pdf>
- Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100-120. <https://doi.org/10.1080/15434303.2019.1674855>
- Kunnan, A. J. (2012). Language assessment for immigration and citizenship. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 162-177). Routledge.
- Llosa, L. (2011). Standards-based classroom assessments of English proficiency: A review of issues, current developments, and future directions for research. *Language Testing*, 28(3), 367-382.
<https://doi.org/10.1177/0265532211404188>
- Llosa, L. (2017). Assessing students' content knowledge and language proficiency. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (3rd ed., pp. 3-15). Springer.
- Ma A. & Norwich B. (2007). Triangulation and theoretical understanding. *International Journal of Social Research Methodology*, 10(3), 211-226.
<https://doi.org/10.1080/13645570701541878>

- Maxwell, J. A. (2008). Designing a qualitative study. In L. Bickman & D. J. Rog (Eds.), *The SAGE handbook of applied social research methods* (pp. 214-253). Springer.
- McNamara, T. (2000). *Language testing*. Oxford University Press.
- Mertler, C. A., & Campbell, C. (2005). *Measuring teachers' knowledge & application of classroom assessment concepts: Development of the assessment literacy inventory* [Paper presentation]. The Annual Meeting of the American Educational Research Association, Montreal, Canada.
<https://files.eric.ed.gov/fulltext/ED490355.pdf>
- Miley, S. K., & Farmer, A. (2017). English language proficiency and content assessment performance: A comparison of English learners and native English speakers' achievement. *English Language Teaching*, 10(9), 198-207.
<https://files.eric.ed.gov/fulltext/EJ1153889.pdf>
- Muhammad, F. H. N. & Bardakçı, M. (2019). Iraqi EFL teachers' assessment literacy: Perceptions and practices. *Arab World English Journal*, 10(2) 420-430. <http://dx.doi.org/10.2139/ssrn.3418574>
- Nasab, F. G. (2015). Alternative versus traditional assessment. *Journal of Applied Linguistics and Language Research*, 2(6), 165-178.
<http://www.jallr.com/index.php/JALLR/article/view/136/pdf136>
- Newfields, T. (2006). Teacher development and assessment literacy. In T. Newfields, I. Gledall, M. Kawate-Mierzejewska, Y. Ishida, M. Chapman, & P. Ross (Eds.), *Authentic communication. Proceedings of the 5th Annual JALT Pan-SIG Conference* (pp. 48-73). Shizuoka, JP: Tokai University College of Marine Science.
<https://hosted.jalt.org/pansig/2006/PDF/Newfields.pdf>

- Ndanu, M. C., & Syombua, M. J. (2015). Mixed methods research: The hidden cracks of the triangulation. *General Education Journal*, 4(2), 46-47.
- Nikmard, F., & Zenouzagh, Z. M. (2020). Designing and validating a potential assessment inventory for assessing ELTs' assessment literacy. *Language Testing in Asia*, 10(1), 1-19. <https://link.springer.com/article/10.1186/s40468-020-00106-1>
- Öz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, 6(1), 25-44.
- Plakans, L. (2015). Integrated second language writing assessment: Why? What? How? *Language and Linguistics Compass*, 9(4), 159-167. <https://doi.org/10.1111/lnc3.12124>
- Plakans, L., Gebril, A., & Bilki, Z. (2019). Shaping a score: Complexity, accuracy, and fluency in integrated writing performances. *Language Testing*, 36(2), 161-179. <https://doi.org/10.1177/0265532216669537>
- Popham, W. J. (2018). *Assessment literacy for educators in a hurry*. ASCD.
- Poza, L. E., & Valdés, G. (2017). Assessing English language proficiency in the United States. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (3rd ed., pp.427-440). Springer.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429-462. <https://doi.org/10.1177/026553220101800407>
- Rea-Dickins, P. (2006). Currents and eddies in the discourse of assessment: A learning focused interpretation. *International Journal of Applied Linguistics*, 16(2), 164-188. <https://doi.org/10.1111/j.1473-4192.2006.00112.x>

- Robles, M., & Braathen, S. (2002). Online assessment techniques. *Delta Pi Epsilon Journal*, 44(1), 39-49. <https://icci.edu.ky/wp-content/uploads/2016/02/online-assessment-techniques.pdf>
- Rus, D. (2019). Assessment techniques in teaching English for specific purposes to engineering students. *Procedia Manufacturing*, 32, 368-373. <https://doi.org/10.1016/j.promfg.2019.02.227>
- Sadeghi, K., & Rahmati, T. (2017). Integrating assessment as, for, and of learning in a large-scale exam preparation course. *Assessing Writing*, 34, 50-61. <https://doi.org/10.1016/j.asw.2017.09.003>
- Sawaki, Y. (2012). Technology in language testing. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 426-437). Routledge.
- Sevimel-Sahin, A. (2021). Language assessment literacy of novice EFL teachers: Perceptions, experiences and training. In S. Hidri (Ed.), *Perspectives on language assessment literacy: Challenges for improved student learning* (pp. 135-158). Routledge.
- Shams, J. A., & Iqbal, M. Z. (2019). Development of classroom assessment literacy training program for university teachers in Punjab. *Bulletin of Education and Research*, 41(2), 41-52. <https://files.eric.ed.gov/fulltext/EJ1229452.pdf>
- Sheean, S. (2017). [Review of the book *Assessment and autonomy in language Learning*, by C. J. Everhard & L. Murphy (Eds.)]. *System*, 70, 121-122. <https://doi.org/10.1016/j.system.2017.06.002>
- Shohamy, E. (1992). Beyond performance testing: A diagnostic feedback testing model for assessing foreign language learning. *Modern Language Journal*, 76(4), 513-521. <https://doi.org/10.2307/330053>

- Smith, M. K. (1996). *What is curriculum? Exploring theory and practice*. Infed.
Retrieved September 12, 2022, from www.infed.org/biblio/b-curric.htm.
- Spolsky, B. (2012). Language testing and language management. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 495-505). Routledge.
- Stanevich, C. (2008). *Building assessment literacy in teachers to promote student achievement* [Unpublished master's thesis]. Prescott College.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-39.
- Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.
<https://doi.org/10.1177/003172170208301010>
- Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324-328.
<https://doi.org/10.1177/003172170508700414>
- Stiggins, R., & Duke, D. (2008). Effective instructional leadership requires assessment leadership. *Phi Delta Kappan*, 90(4), 285-291.
<https://doi.org/10.1177/003172170809000410>
- Stiggins, R. (2014). Improve assessment literacy outside of schools too. *Phi Delta Kappan* 96(2), 67–72. <https://doi.org/10.1177/0031721714553413>
- Şişman, E. P., & 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>
- Taber, K. S. (2018). The use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48, 1273–1296. <https://link.springer.com/article/10.1007/s11165-016-9602-2>

- Talib, R., Kamsah, M. Z., Ghafar, M. N. A., Zakaria, M. A. Z. M., & Naim, H. A. (2013). T-Assess: Assessment literacy for Malaysian teachers. *Proceedings of the International Conference on Assessment for Higher Education Across Domains and Skills*, Kuala Lumpur.
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36. <https://doi.org/10.1017/S0267190509090035>
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412. <https://doi.org/10.1177/0265532213480338>
- Trotter, E. (2006). Student perceptions of continuous summative assessment. *Assessment & Evaluation in Higher Education*, 31(5), 505-521. <https://doi.org/10.1080/02602930600679506>
- Tsagari, D. (2011). Washback of a high-stakes English exam on teachers' perceptions and practices. *Selected Papers on Theoretical and Applied Linguistics*, 19, 431-445. <https://doi.org/10.26262/istal.v19i0.5521>
- Tsagari, D., & Cheng, L. (2017). Washback, impact, and consequences revisited. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (3rd ed., pp. 359-372). Springer.
- Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, 6(1), 41-63. https://web.archive.org/web/20180412230835id_/https://arts.unimelb.edu.au/data/assets/pdf_file/0010/2349928/6_1_SI3TsagariVogt.pdf

- Tsagari, D. (2021). Language assessment literacy: Concepts, challenges and prospects. In S. Hidri (Ed.), *Perspectives on language assessment literacy: Challenges for improved student learning* (pp. 31-32). Routledge.
- Turner, S., Zanker, N., & Braine, M. (2012). An investigation into teacher wellbeing during the teacher training year. *Design and Technology Education: An International Journal*, 17(2), 21-34.
<http://ojs.lboro.ac.uk/ojs/index.php/DATE/article/view/1729>
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374-402. <https://doi.org/10.1080/15434303.2014.960046>
- Vogt, K., Tsagari, D., & Spanoudis, G. (2020). What do teachers think they want? A comparative study of in-service language teachers' beliefs on LAL training needs. *Language Assessment Quarterly*, 17(4), 386-409.
<https://doi.org/10.1080/15434303.2020.1781128>
- Wall, D. (2012). Washback. In G. Fulcher & F. Davidson, (Eds.), *The Routledge handbook of language testing* (pp. 79-92). Routledge.
- Wallace, M. P., & Qin, C. Y. (2021). Language classroom assessment fairness: Perceptions from students. *LEARN Journal: Language Education and Acquisition Research Network*, 14(1), 492-521.
<https://files.eric.ed.gov/fulltext/EJ1284499.pdf>
- Wiliam, D., & Thompson, M. (2017). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment* (pp. 53-82). Routledge.
- Westhuizen, D. (2016). *Guidelines for online assessment for educators*. Commonwealth of Learning.

- Wolf, M. K., & Faulkner-Bond, M. (2016). Validating English language proficiency assessment uses for English learners: Academic language proficiency and content assessment performance. *Educational Measurement: Issues and Practice*, 35(2), 6-18. <https://doi.org/10.1111/emip.12105>
- Wrigley, S. H. (1992). Assessing ongoing progress: Are we progressing? In D. D. Holt (Ed.), *Alternative approaches to assessment & evaluation in family English literacy programs* (pp. 57-84). California Department of Education.
- Xu, Y., & Brown, G. T. L. (2017). University English teacher assessment literacy: A survey test report from China. *Papers in Language Testing and Assessment*, 6(1), 133-158. https://arts.unimelb.edu.au/data/assets/pdf_file/0011/2349947/6_1_SI7XuBrown.pdf
- Yildirim, R., & Orsdemir, E. (2013). Performance tasks as alternative assessment for young EFL learners: Does practice match the curriculum proposal? *International Online Journal of Educational Sciences*, 5(3), 562-574. <http://www.ajindex.com/dosyalar/makale/acarindex-1423904198.pdf>
- Yıldırım, A. & Şimşek, H. (2016). *Sosyal bilimlerde nitel araştırma yöntemleri*. Seçkin.
- Yin, R. K. (1981). The case study crisis: Some answers. *Administrative Science Quarterly*, 26(1), 58-65. <https://doi.org/10.2307/2392599>
- Yin, R. K. (2003). Designing case studies. In L. Maruster & M. J. Gijsenberg (Eds.), *Qualitative Research Methods* (pp. 359-386). Sage Publications.
- Yu, G. (2013). From integrative to integrated language assessment: Are we there yet? *Language Assessment Quarterly*, 10(1), 110-114. <https://doi.org/10.1080/15434303.2013.766744>

Zulaiha, S., Mulyono, H., & Ambarsari, L. (2020). An investigation into EFL teachers' assessment literacy: Indonesian teachers' perceptions and classroom practice. *European Journal of Contemporary Education*, 9(1), 189-201.

<https://files.eric.ed.gov/fulltext/EJ1249376.pdf>



APPENDICES

Appendix A

Assessment Literacy Survey (Teachers)

Dear Teacher,

This survey* was prepared for my thesis, within the scope of İhsan Doğramacı Bilkent University, Curriculum and Instruction Master's Program, which aims to explore the perceived relationship between assessment literacy of English language professionals and their teaching quality under the supervision of Asst. Prof. Dr. Tijen Akşit. The survey will take approximately 15 minutes to complete. In terms of this survey;

- Participation in this study is voluntary
- Your identification will be kept confidential.
- Your data will only be accessed by the researcher, will be kept on a password protected computer and will only be used for this thesis and possible related conference presentations, and journal publication purposes.
- Your answers will not be shared with your institution and will not affect your working rights.
- You can leave the survey any time you want.
- The results of the study will be shared with you if you request them. In this case, you can reach me via the e-mail address stated below.

Please tick the box below if you meet the criteria and agree to participate.

Criteria:

- I am over 18 years old.
- I have read and understood the information about this study above.
- I understand that I can withdraw from the study without any consequences at any time.
- I understand who would have access to identifying information provided and what will happen to the data at the end of the study.
- I understand that this study has been reviewed by and received ethical clearance through Bilkent University Ethics Committee.

*The survey was adapted from Kremmel B. & Harding, L. (2020)

I meet the criteria and agree to participate.

☐

Pınar KILIÇ

Bilkent University/ MA in CI

e-mail:

Advisor: Asst. Prof. Dr. Tijen AKŞİT

PART A: BACKGROUND**1. Gender**Female ☐Male ☐Prefer not to say ☐**2. Age**21-25 ☐26-30 ☐31-35 ☐36-40 ☐40+ ☐**3. Years of experience in teaching English at this institution**0-5 years ☐6-10 years ☐11-15 years ☐16-20 years ☐**4. Years of experience in teaching English in total**0-5 years ☐6-10 years ☐11-15 years ☐16-20 years ☐21+years ☐**5. Level of education**Bachelor Degree ☐MA/MSc Student ☐MA/MSc Degree ☐PhD student ☐PhD Degree ☐

6. Department

Bachelor Degree ☐

English Language Teaching ☐

American Culture and Literature ☐

English Language and Literature ☐

English Translation and Interpreting ☐

Linguistics ☐

Other Please specify _____ ☐

Master's Degree

English Language Teaching ☐

American Culture and Literature ☐

English Language and Literature ☐

English Translation and Interpreting ☐

Linguistics ☐

Curriculum and Instruction ☐

Measurement and Evaluation ☐

Other Please specify _____ ☐

PhD Degree

English Language Teaching ☐

American Culture and Literature ☐

English Language and Literature ☐

English Translation and Interpreting ☐

Linguistics ☐

Curriculum and Instruction ☐

Measurement and Evaluation ☐

Other Please specify _____ ☐

7. Did you have any additional responsibilities at school before? (i.e. units and administration)Yes ☐No ☐**Please tick the one that applies:**

I was a member of Curriculum Development Unit.	☐
I was a member of Measurement and Evaluation Unit.	☐
I was a member of Administrative Unit.	☐
I was a member of Educational Technologies Unit.	☐
I was a member of Extracurricular Activities Unit.	☐

8. Did you attend the sessions of Peer-Learning: Assessment course?Yes ☐No ☐

PART B**I. How knowledgeable do you think you are about each aspect of language assessment below? Please respond according to the following scale:**

	Not knowledgeable at all 1	Slightly knowledgeable 2	Moderately knowledgeable 3	Very knowledgeable 4	Extremely knowledgeable 5
1.	how to use assessments to inform learning or teaching goals				
2.	how to use assessments to evaluate achievement in language learning				
3.	how to use assessments to diagnose learners' strengths and weaknesses				
4.	how to use assessments to motivate student learning				
5.	how to use self-assessment				
6.	how to use peer-assessment				
7.	how to interpret what a particular score says about an individual's language ability				
8.	how to determine if a language assessment aligns with local systems of accreditation (YÖKAK and DEDAK)				
9.	how to determine if a language assessment aligns with YÖK (Council of Higher Education) system.				
10.	how to determine if the content of a language assessment is culturally appropriate				
11.	how to determine if the results from a language assessment are relevant to the context of Turkish preparatory schools				

12.	how to train others about language assessment	1	2	3	4	5
13.	how to recognize when an assessment is being used inappropriately	1	2	3	4	5
14.	how to prepare learners to take language assessments	1	2	3	4	5
15.	how to give useful feedback on the basis of an assessment	1	2	3	4	5
16.	how assessments can be used to enforce social policies (e.g., immigration)	1	2	3	4	5
17.	how assessments can influence teaching and learning in the classroom	1	2	3	4	5
18.	how assessments can influence teaching and learning materials	1	2	3	4	5
19.	how assessments can influence the design of a language course or curriculum	1	2	3	4	5
20.	how language skills develop (e.g., reading, listening, writing, speaking)	1	2	3	4	5
21.	how foreign/second languages are learned	1	2	3	4	5
22.	how language is used in society	1	2	3	4	5
23.	how social values can influence language assessment design and use	1	2	3	4	5
24.	the concept of reliability (how accurate or consistent an assessment is)	1	2	3	4	5
25.	the concept of validity (how well an assessment measures what it claims to measure)	1	2	3	4	5
26.	the structure of language	1	2	3	4	5

27.	the relevant legal regulations for assessment in Turkish preparatory schools	1	2	3	4	5
28.	the assessment traditions in Turkish preparatory schools	1	2	3	4	5
29.	different types of purposes for language assessment purposes (e.g., proficiency, achievement, diagnostic)	1	2	3	4	5
30.	one's own beliefs/attitudes towards language assessment	1	2	3	4	5
31.	how one's own beliefs/attitudes might influence one's assessment practices	1	2	3	4	5
32.	how one's own beliefs/attitudes may conflict with those of other groups involved in assessment	1	2	3	4	5
33.	how one's own knowledge of language assessment might be further developed	1	2	3	4	5

**II. How skilled do you think you are in each aspect of language assessment below?
Please respond according to the following scale:**

Not skilled at all	Slightly skilled	Moderately skilled	Very skilled	Extremely skilled
1	2	3	4	5

33.	using statistics to analyse the difficulty of individual items (questions) or tasks	1	2	3	4	5
34.	using statistics to analyse overall scores on a particular assessment	1	2	3	4	5
35.	using statistics to analyse the quality of individual items/tasks	1	2	3	4	5

36.	using techniques other than statistics (e.g., questionnaires, interviews, analysis of language) to get information about the quality of a language assessment	1	2	3	4	5
37.	using rating scales (rubrics) to score speaking or writing performances	1	2	3	4	5
38.	scoring closed-response questions (e.g. Multiple Choice Questions)	1	2	3	4	5
39.	scoring open-ended questions (e.g. short answer questions)	1	2	3	4	5
40.	developing specifications (overall plans) for language assessments	1	2	3	4	5
41.	selecting appropriate rating scales (rubrics)	1	2	3	4	5
42.	selecting appropriate items or tasks for a particular assessment purpose	1	2	3	4	5
43.	training others to use rating scales (rubrics) appropriately	1	2	3	4	5
44.	training others to write good quality items (questions) or tasks for language assessments	1	2	3	4	5
45.	writing good quality items (questions) or tasks for language assessments	1	2	3	4	5
46.	aligning tests to proficiency frameworks (e.g., the Common European Framework of Reference)	1	2	3	4	5
47.	determining pass-fail marks/cut-scores	1	2	3	4	5
48.	identifying assessment bias	1	2	3	4	5
49.	accommodating candidates with disabilities or other learning impairments	1	2	3	4	5

50.	designing scoring keys and rating scales (rubrics) for assessment tasks	1	2	3	4	5
51.	making decisions about what aspects of language to assess	1	2	3	4	5
52.	piloting/trying-out assessments before their administration	1	2	3	4	5

IMPORTANT NOTE: If you would like to volunteer for a short interview to provide me with further information, please share your e-mail address and/or phone number below.

e-mail: _____ mobile: +90 (5____) _____

Thank you for your contribution.



Appendix B

Assessment Literacy Survey (Assessment Specialists)

Dear Assessment Specialist,

This survey* was prepared for my thesis, within the scope of İhsan Doğramacı Bilkent University, Curriculum and Instruction Master's Program, which aims to explore the perceived relationship between assessment literacy of English language professionals and their teaching quality under the supervision of Asst. Prof. Dr. Tijen Akşit. The survey will take approximately 15 minutes to complete. In terms of this survey;

- Participation in this study is voluntary
- Your identification will be kept confidential.
- Your data will only be accessed by the researcher, will be kept on a password protected computer and will only be used for this thesis and possible related conference presentations, and journal publication purposes.
- Your answers will not be shared with your institution and will not affect your working rights.
- You can leave the survey any time you want.
- The results of the study will be shared with you if you request them. In this case, you can reach me via the e-mail address stated below.

Please tick the box below if you meet the criteria and agree to participate.

Criteria:

- I am over 18 years old.
- I have read and understood the information about this study above.
- I understand that I can withdraw from the study without any consequences at any time.
- I understand who would have access to identifying information provided and what will happen to the data at the end of the study.
- I understand that this study has been reviewed by and received ethical clearance through Bilkent University Ethics Committee.

*The survey was adapted from Kremmel B. & Harding, L. (2020)

I meet the criteria and agree to participate.

☐

Pınar KILIÇ

Bilkent University/ MA in CI

e-mail:

Advisor: Asst. Prof. Dr. Tijen AKŞİT

PART A: BACKGROUND INFORMATION**1. Gender**Female ☐Male ☐Prefer not to say ☐**2. How old are you?**21-25 ☐26-30 ☐31-35 ☐36-40 ☐40+ ☐**3. Years of experience in teaching English at this institution**0-5 years ☐6-10 years ☐11-15 years ☐16-20 years ☐**4. Years of experience in teaching English in total**0-5 years ☐6-10 years ☐11-15 years ☐16-20 years ☐21+years ☐**5. Years of experience as an assessment specialist in the Measurement and Evaluation Unit at this institution**0-2 years ☐3-5 years ☐6-8 years ☐9-11 years ☐12+ years ☐

6. Years of experience as an assessment specialist in total0-2 years ☐3-5 years ☐6-8 years ☐9-11 years ☐**12 years** ☐**7. Level of education**Bachelor Degree ☐MA/MSc Student ☐MA/MSc Degree ☐PhD student ☐PhD Degree ☐**8. Department**

Bachelor Degree

English Language Teaching ☐American Culture and Literature ☐English Language and Literature ☐English Translation and Interpreting ☐Linguistics ☐Other ☐ Please specify _____**Master's Degree**English Language Teaching ☐American Culture and Literature ☐English Language and Literature ☐English Translation and Interpreting ☐Linguistics ☐Curriculum and Instruction ☐Measurement and Evaluation ☐Other ☐ Please specify _____**PhD Degree**English Language Teaching ☐American Culture and Literature ☐English Language and Literature ☐English Translation and Interpreting ☐Linguistics ☐Curriculum and Instruction ☐Measurement and Evaluation ☐Other ☐ Please specify _____

PART B

- I. How knowledgeable do you think “the instructors in your school” are about each aspect of language assessment below? Please respond according to the following scale:**

	Not knowledgeable at all 1	Slightly knowledgeable 2	Moderately knowledgeable 3	Very knowledgeable 4	Extremely knowledgeable 5
1.	how to use assessments to inform learning or teaching goals				
2.	how to use assessments to evaluate achievement in language learning				
3.	how to use assessments to diagnose learners’ strengths and weaknesses				
4.	how to use assessments to motivate student learning				
5.	how to use self-assessment				
6.	how to use peer-assessment				
7.	how to interpret what a particular score says about an individual’s language ability				
8.	how to determine if a language assessment aligns with local systems of accreditation (YÖKAK and DEDAK)				
9.	how to determine if a language assessment aligns with YÖK (Council of Higher Education) system.				
10.	how to determine if the content of a language assessment is culturally appropriate				
11.	how to determine if the results from a language assessment are relevant to the context of Turkish preparatory schools				

12.	how to train others about language assessment	1	2	3	4	5
13.	how to recognize when an assessment is being used inappropriately	1	2	3	4	5
14.	how to prepare learners to take language assessments	1	2	3	4	5
15.	how to give useful feedback on the basis of an assessment	1	2	3	4	5
16.	how assessments can be used to enforce social policies (e.g., immigration)	1	2	3	4	5
17.	how assessments can influence teaching and learning in the classroom	1	2	3	4	5
18.	how assessments can influence teaching and learning materials	1	2	3	4	5
19.	how assessments can influence the design of a language course or curriculum	1	2	3	4	5
20.	how language skills develop (e.g., reading, listening, writing, speaking)	1	2	3	4	5
21.	how foreign/second languages are learned	1	2	3	4	5
22.	how language is used in society	1	2	3	4	5
23.	how social values can influence language assessment design and use	1	2	3	4	5
24.	the concept of reliability (how accurate or consistent an assessment is)	1	2	3	4	5
25.	the concept of validity (how well an assessment measures what it claims to measure)	1	2	3	4	5
26.	the structure of language	1	2	3	4	5

27.	the relevant legal regulations for assessment in Turkish preparatory schools	1	2	3	4	5
28.	the assessment traditions in Turkish preparatory schools	1	2	3	4	5
29.	different types of purposes for language assessment purposes (e.g., proficiency, achievement, diagnostic)	1	2	3	4	5
30.	one's own beliefs/attitudes towards language assessment	1	2	3	4	5
31.	how one's own beliefs/attitudes might influence one's assessment practices	1	2	3	4	5
32.	how one's own beliefs/attitudes may conflict with those of other groups involved in assessment	1	2	3	4	5
33.	how one's own knowledge of language assessment might be further developed	1	2	3	4	5

II. How skilled do you think “the instructors in your school” are about each aspect of language assessment below? Please respond according to the following scale:

Not skilled
at all

1

Slightly
skilled

2

Moderately
skilled

3

Very skilled

4

Extremely
skilled

5

33.	using statistics to analyse the difficulty of individual items (questions) or tasks	1	2	3	4	5
34.	using statistics to analyse overall scores on a particular assessment	1	2	3	4	5
35.	using statistics to analyse the quality of individual items/tasks	1	2	3	4	5

36.	using techniques other than statistics (e.g., questionnaires, interviews, analysis of language) to get information about the quality of a language assessment	1	2	3	4	5
37.	using rating scales (rubrics) to score speaking or writing performances	1	2	3	4	5
38.	scoring closed-response questions (e.g. Multiple Choice Questions)	1	2	3	4	5
39.	scoring open-ended questions (e.g. short answer questions)	1	2	3	4	5
40.	developing specifications (overall plans) for language assessments	1	2	3	4	5
41.	selecting appropriate rating scales (rubrics)	1	2	3	4	5
42.	selecting appropriate items or tasks for a particular assessment purpose	1	2	3	4	5
43.	training others to use rating scales (rubrics) appropriately	1	2	3	4	5
44.	training others to write good quality items (questions) or tasks for language assessments	1	2	3	4	5
45.	writing good quality items (questions) or tasks for language assessments	1	2	3	4	5
46.	aligning tests to proficiency frameworks (e.g., the Common European Framework of Reference)	1	2	3	4	5
47.	determining pass-fail marks/cut-scores	1	2	3	4	5
48.	identifying assessment bias	1	2	3	4	5
49.	accommodating candidates with disabilities or other learning impairments	1	2	3	4	5

50.	designing scoring keys and rating scales (rubrics) for assessment tasks	1	2	3	4	5
51.	making decisions about what aspects of language to assess	1	2	3	4	5
52.	piloting/trying-out assessments before their administration	1	2	3	4	5

IMPORTANT NOTE: If you would like to volunteer for a short interview to provide me with further information, please share your e-mail address and/or phone number below.

e-mail: _____ mobile: +90 (5____) _____

Thank you for your contribution.



Appendix C

Assessment Literacy Survey (Academic Leaders)

Dear Academic Leader,

This survey* was prepared for my thesis, within the scope of İhsan Doğramacı Bilkent University, Curriculum and Instruction Master's Program, which aims to explore the perceived relationship between assessment literacy of English language professionals and their teaching quality under the supervision of Asst. Prof. Dr. Tijen Akşit. The survey will take approximately 15 minutes to complete. In terms of this survey;

- Participation in this study is voluntary
- Your identification will be kept confidential.
- Your data will only be accessed by the researcher, will be kept on a password protected computer and will only be used for this thesis and possible related conference presentations, and journal publication purposes.
- Your answers will not be shared with your institution and will not affect your working rights.
- You can leave the survey any time you want.
- The results of the study will be shared with you if you request them. In this case, you can reach me via the e-mail address stated below.

Please tick the box below if you meet the criteria and agree to participate.

Criteria:

- I am over 18 years old.
- I have read and understood the information about this study above.
- I understand that I can withdraw from the study without any consequences at any time.
- I understand who would have access to identifying information provided and what will happen to the data at the end of the study.
- I understand that this study has been reviewed by and received ethical clearance through Bilkent University Ethics Committee.

*The survey was adapted from Kremmel B. & Harding, L. (2020)

I meet the criteria and agree to participate.

☐

Pınar KILIÇ

Bilkent University/ MA in CI

e-mail

Advisor: Asst. Prof. Dr. Tijen AKŞİT

PART A: BACKGROUND INFORMATION**9. Gender**Female ☐Male ☐Prefer not to say ☐**10. Age**21-25 ☐26-30 ☐31-35 ☐36-40 ☐40+ ☐**11. Years of experience in teaching English at this institution**0-5 years ☐6-10 years ☐11-15 years ☐16-20 years ☐**12. Years of experience in teaching English in total**0-5 years ☐6-10 years ☐11-15 years ☐16-20 years ☐21+years ☐**13. Years of experience as an academic leader at this institution**0-2 years ☐3-5 years ☐6-8 years ☐9-11 years ☐12-14 years ☐**14. Years of experience as an academic leader in total**0-2 years ☐3-5 years ☐6-8 years ☐

- 9-11 years ☐
- 12-14 years ☐
- 14+ years ☐

15. Level of education

- Bachelor Degree ☐
- MA/MSc Student ☐
- MA/MSc Degree ☐
- PhD student ☐
- PhD Degree ☐

16. Department

- Bachelor Degree
- English Language Teaching ☐
- American Culture and Literature ☐
- English Language and Literature ☐
- English Translation and Interpreting ☐
- Linguistics ☐
- Other ☐ Please specify _____

Master's Degree

- English Language Teaching ☐
- American Culture and Literature ☐
- English Language and Literature ☐
- English Translation and Interpreting ☐
- Linguistics ☐
- Curriculum and Instruction ☐
- Measurement and Evaluation ☐
- Other ☐ Please specify _____

PhD Degree

- English Language Teaching ☐
- American Culture and Literature ☐
- English Language and Literature ☐
- English Translation and Interpreting ☐
- Linguistics ☐
- Curriculum and Instruction ☐
- Measurement and Evaluation ☐
- Other ☐ Please specify _____

PART B

- I. How knowledgeable do you think the “instructors in your school (excluding assessment specialists)” are about each aspect of language assessment below? Please respond according to the following scale:**

	Not knowledgeable at all 1	Slightly knowledgeable 2	Moderately knowledgeable 3	Very knowledgeable 4	Extremely knowledgeable 5
1.	how to use assessments to inform learning or teaching goals				
2.	how to use assessments to evaluate achievement in language learning				
3.	how to use assessments to diagnose learners' strengths and weaknesses				
4.	how to use assessments to motivate student learning				
5.	how to use self-assessment				
6.	how to use peer-assessment				
7.	how to interpret what a particular score says about an individual's language ability				
8.	how to determine if a language assessment aligns with local systems of accreditation (YÖKAK and DEDAK)				
9.	how to determine if a language assessment aligns with YÖK (Council of Higher Education) system.				
10.	how to determine if the content of a language assessment is culturally appropriate				
11.	how to determine if the results from a language assessment are relevant to the context of Turkish preparatory schools				
12.	how to train others about language assessment				

13.	how to recognize when an assessment is being used inappropriately	1	2	3	4	5
14.	how to prepare learners to take language assessments	1	2	3	4	5
15.	how to give useful feedback on the basis of an assessment	1	2	3	4	5
16.	how assessments can be used to enforce social policies (e.g., immigration)	1	2	3	4	5
17.	how assessments can influence teaching and learning in the classroom	1	2	3	4	5
18.	how assessments can influence teaching and learning materials	1	2	3	4	5
19.	how assessments can influence the design of a language course or curriculum	1	2	3	4	5
20.	how language skills develop (e.g., reading, listening, writing, speaking)	1	2	3	4	5
21.	how foreign/second languages are learned	1	2	3	4	5
22.	how language is used in society	1	2	3	4	5
23.	how social values can influence language assessment design and use	1	2	3	4	5
24.	the concept of reliability (how accurate or consistent an assessment is)	1	2	3	4	5
25.	the concept of validity (how well an assessment measures what it claims to measure)	1	2	3	4	5
26.	the structure of language	1	2	3	4	5
27.	the relevant legal regulations for assessment in Turkish preparatory schools	1	2	3	4	5
28.	the assessment traditions in Turkish preparatory schools	1	2	3	4	5

29.	different types of purposes for language assessment purposes (e.g., proficiency, achievement, diagnostic)	1	2	3	4	5
30.	one's own beliefs/attitudes towards language assessment	1	2	3	4	5
31.	how one's own beliefs/attitudes might influence one's assessment practices	1	2	3	4	5
32.	how one's own beliefs/attitudes may conflict with those of other groups involved in assessment	1	2	3	4	5
33.	how one's own knowledge of language assessment might be further developed	1	2	3	4	5

**II. How skilled do you think the “instructors in your school (excluding assessment specialists)” are in each aspect of language assessment below?
Please respond according to the following scale:**

Not skilled
at all

1

Slightly
skilled

2

Moderately
skilled

3

Very skilled

4

Extremely
skilled

5

33.	using statistics to analyse the difficulty of individual items (questions) or tasks	1	2	3	4	5
34.	using statistics to analyse overall scores on a particular assessment	1	2	3	4	5
35.	using statistics to analyse the quality of individual items/tasks	1	2	3	4	5
36.	using techniques other than statistics (e.g., questionnaires, interviews, analysis of language) to get information about the quality of a language assessment	1	2	3	4	5
37.	using rating scales (rubrics) to score speaking or writing performances	1	2	3	4	5
38.	scoring closed-response questions (e.g. Multiple Choice Questions)	1	2	3	4	5

39.	scoring open-ended questions (e.g. short answer questions)	1	2	3	4	5
40.	developing specifications (overall plans) for language assessments	1	2	3	4	5
41.	selecting appropriate rating scales (rubrics)	1	2	3	4	5
42.	selecting appropriate items or tasks for a particular assessment purpose	1	2	3	4	5
43.	training others to use rating scales (rubrics) appropriately	1	2	3	4	5
44.	training others to write good quality items (questions) or tasks for language assessments	1	2	3	4	5
45.	writing good quality items (questions) or tasks for language assessments	1	2	3	4	5
46.	aligning tests to proficiency frameworks (e.g., the Common European Framework of Reference)	1	2	3	4	5
47.	determining pass-fail marks/cut-scores	1	2	3	4	5
48.	identifying assessment bias	1	2	3	4	5
49.	accommodating candidates with disabilities or other learning impairments	1	2	3	4	5
50.	designing scoring keys and rating scales (rubrics) for assessment tasks	1	2	3	4	5
51.	making decisions about what aspects of language to assess	1	2	3	4	5
52.	piloting/trying-out assessments before their administration	1	2	3	4	5

IMPORTANT NOTE: If you would like to volunteer for a short interview to provide me with further information, please share your e-mail address and/or phone number below.

e-mail: _____ mobile: +90 (5__) _____

Thank you for your contribution.

Appendix D

Interview Questions for English Language Professionals

Dear Instructor,

This interview was designed for my thesis within the scope of İhsan Doğramacı Bilkent University, Curriculum and Instruction Master's Program. It aims to explore the perceived relationship between assessment literacy of English language professionals and their teaching quality under the supervision of Asst. Prof. Dr. Tijen Akşit. The interview will take approximately 20 minutes. In terms of this interview;

- Participation in this study is voluntary.
- Your voice will be recorded.
- Your identification will be kept confidential.
- Your data will only be accessed by the researcher, will be kept on a password protected computer and will only be used for this thesis and possible related conference presentations, and journal publication purposes.
- Your answers will not be shared with your institution and will not affect your working rights.
- You can leave the interview any time you want.
- Your questions related to the study and the interview will be answered.
- The results of the study will be shared with you if you request them. In this case, you can reach me via the e-mail address stated below.

Pınar KILIÇ

Bilkent University/ MA in CI

e-mail:

Advisor: Asst. Prof. Dr. Tijen AKŞİT

I have read the information above and agree to participate in the interview.

(Please sign and submit the form to the researcher).

Name and Surname

Date

Signature

---/---/---

1. How would you describe the student assessment system (*components in the assessment schedule*) at your institution?

Follow-up: What do you think about these assessment practices?

Follow-up: What kind of assessment tools do you prefer in your classroom?

2. How do you see the role of student assessment for an English language instructor/ assessment specialist/ an academic leader?

- *in class*
- *at your institution*
- *at preparatory schools in Turkey*

3. How important do you think teachers' competency in language assessment is for teaching quality?

Follow-up: Can you think of some specific situations where your competency in language assessment has contributed to your teaching quality?

4. What do you think about your assessment literacy level? Please explain and/or exemplify why you think in that way.

Follow-up: If you feel the need, how could you improve it?

5. Is there anything else you would like to share about this topic?

Appendix E

Indication of Coding Categories

Main Categories	Sub-Categories	Concepts	Codes
KNOWLEDGE	Knowledge of Theory	<i>Structure of English</i> <i>Learning English</i>	K-KT-SE K-KT-LE
	Language Pedagogy	<i>Student Motivation</i> <i>Washback</i> <i>Feedback</i> <i>Student Achievement</i> <i>In-class Assessments</i>	K-LP-SM K-LP-W K-LP-F K-LP-SA K-LP-IA
	Principles and Concepts	<i>Validity</i> <i>Reliability</i> <i>Types of Assessment</i>	K-PC-V K-PC-R K-PC-TA
	Impact of Social Values	<i>Social Values</i>	K-ISV-SV
	Local Practices	<i>Accreditation</i> <i>Turkish Assessment Culture</i> <i>Training Others</i>	K-LoP-A K-LoP-TAC K-LoP-TO
	Personal Beliefs & Attitudes	<i>Personal Beliefs</i> <i>How to Improve</i>	K-PBA-PB K-PBA-HI
	Scores and Decision Making	<i>Score Interpretation</i> <i>Decision Making</i>	K-SDM-DM K-SDM-DM
	*Overall Assessment Literacy Level	*Overall Knowledge Level	*K- OAL-OKL
SKILLS	Constructing Language Assessments	<i>Item Writing</i> <i>Designing Rubrics</i> <i>Test Specifications</i>	S-CLA-IW S-CLA-DR S-CLA-TS
	Administering and Scoring Language Assessments	<i>Scoring Items</i> <i>Using Rubrics</i>	S-ASLA-SI S-ASLA-UR
	Evaluating Language Assessments	<i>Statistics</i> <i>Proficiency Frameworks</i>	S-ELA-S S-ELA-PF
	*Overall Assessment Literacy Level	*Overall Skills Level	*K- OAL-OSL
	*Case Descriptions	<i>*Assessment Practices</i> <i>*Proficiency Exam</i> <i>*Recommendations</i>	*CD-AP *CD-TOEFL *CD-R