

**ASSESSMENT PRACTICES
IN THE PREPARATORY LANGUAGE SCHOOLS
IN TURKISH HIGHER EDUCATION
INSTITUTIONS:
PROBLEMS AND SOLUTIONS**

PhD Dissertation

Mehmet ŞENGÜL

Eskişehir 2022

**ASSESSMENT PRACTICES IN THE PREPARATORY LANGUAGE SCHOOLS IN
TURKISH HIGHER EDUCATION INSTITUTIONS:
PROBLEMS AND SOLUTIONS**

Mehmet ŞENGÜL

PhD DISSERTATION

**Department of Foreign Language Education
Programme in English Language Teaching
Supervisor: Assoc. Prof. Dr. Ali MERCİ**

**Eskişehir
Anadolu University
Graduate School of Educational Sciences
December 2022**

JÜRİ VE ENSTİTÜ ONAYI

ABSTRACT

ASSESSMENT PRACTICES IN THE PREPARATORY LANGUAGE SCHOOLS IN TURKISH HIGHER EDUCATION INSTITUTIONS: PROBLEMS AND SOLUTIONS

Mehmet ŞENGÜL

Department of Foreign Language Education
Programme in English Language Teaching
Anadolu University, Graduate School of Educational Sciences, December 2022

Supervisor: Assoc. Prof. Dr. Ali MERÇ

In Türkiye, students enrolled in universities study commonly English during their first year at English Preparatory Programmes (EPPs). Due to washback effect, language assessment influences different practices ranging from learning-teaching to learners' or teachers' attitudes. Therefore, language assessment plays a crucial role in EPPs. In several schools, there are testing units responsible for providing exams. The literature review shows that several challenges are encountered in assessment practices remaining as a dead-end. Therefore, this study aimed to provide an in-depth analysis of the challenges of the testing processes and produce solutions.

This is a qualitative study in two phases. In the first phase, the researcher attempted to identify the challenges through the interviews with the directors and TUMs from 50 universities. In the second phase, it was aimed at finding solutions employing the Delphi method with three rounds. According to the findings, four major categories emerged, namely systemic, stakeholder-related, testing processes-related, and general challenges. The findings reveal that EPPs do not implement a standard procedure due to several reasons such as lack of staff, experts, and time. In addition, it was found that the instructors working at TUs are reluctant to work because of extra workload, no extra payment or incentives. To conclude, it is crucial this issue must be taken into the agenda of high-level authorities. At the end of the study, implications to different stakeholders including YÖK, MEB, ÖSYM, researchers, and TUMs were provided.

Keywords: Higher education, Foreign language education, Language testing and assessment, Challenges and solutions

ÖZET

TÜRK YÜKSEKÖĞRETİMİNDE HAZIRLIK OKULLARINDA DEĞERLENDİRME UYGULAMALARI: ZORLUKLAR VE ÇÖZÜM ÖNERİLERİ

Mehmet ŞENGÜL

Yabancı Diller Eğitimi Anabilim Dalı
İngilizce Öğretmenliği Programı
Anadolu Üniversitesi, Eğitim Bilimleri Enstitüsü, Aralık 2022

Danışman: Doç. Dr. Ali MERÇ

Türkiye'de üniversitelere kayıt yaptıran öğrenciler, ilk yıllarında, İngilizce Hazırlık Programlarında (EPP'ler) yaygın olarak İngilizce öğrenmektedir. Sınav etkisi nedeniyle, yabancı dil değerlendirmesi, öğrenme-öğretmeden öğrencilerin veya öğretmenlerin tutumlarına kadar farklı uygulamaları etkilemektedir. Bu nedenle, yabancı dil değerlendirmesi EPP'lerde çok önemli bir rol oynamaktadır. Birkaç okulda, sınavların hazırlanmasından sorumlu sınav birimleri vardır. Literatür taraması, değerlendirme uygulamalarında çıkmaz sokak olarak süregelen çeşitli zorlukların yaşandığını göstermektedir. Bu nedenle bu çalışma, test süreçlerindeki zorlukların ve çözümlerin derinlemesine analizini sunmayı amaçlamaktadır.

Bu, iki aşamalı nitel bir çalışmadır. İlk aşamada araştırmacı, 50 üniversitedeki müdürler ve sınav hazırlayanlar ile yapılan görüşmeler ile zorlukları belirlemeye çalışmıştır. İkinci aşamada ise, üç turlu Delphi yöntemi kullanılarak çözümlerin bulunması hedeflenmiştir. Bulgulara göre, sistemik, paydaşlar ve test süreçleri ile ilgili ve genel zorluklar olarak dört ana kategori ortaya çıkmıştır. Bulgular, EPP'lerde personel, uzman ve zaman eksikliği gibi birçok nedenden dolayı standart bir prosedürün uygulanmadığını ortaya koymaktadır. Ayrıca, sınav birimi üyelerinin fazla iş yükü olması, ek ödeme veya teşviklerin olmaması nedenleriyle isteksiz oldukları bulunmuştur. Sonuç olarak, konunun üst düzey yetkililerin gündemine alınması önemlidir. Çalışmanın sonunda, YÖK, MEB, ÖSYM, araştırmacılar ve sınav birimi çalışanları dahil olmak üzere farklı paydaşlara öneriler getirilmiştir.

Anahtar Sözcükler: Yükseköğretim, Yabancı dil eğitimi, Yabancı dilde ölçme-değerlendirme, Zorluklar ve çözümler önerileri.

ACKNOWLEDGEMENT

I have worked with a significant number of people during the preparation and writing process of my thesis. Many people, including my friends, made crucial contributions with their help, so I am pleased to convey my gratitude to them.

First of all, I would like to express my most profound appreciation to my supervisor, Assoc. Prof. Dr. Ali MERÇ, who is there and ready when needed. His endless support, guidance, criticism, advice, and encouragement, along with his timeless patience during my thesis preparation and writing process, were invaluable to me. Furthermore, he shared his wisdom and experience with me during the study.

I want to extend my respect and gratitude to the members of the core examining committee: Prof. Dr. Özgür YILDIRIM and Assist. Prof. Dr. Sercan SAĞLAM, both of whom gave support and suggestion for my work, which turned this study into a fruitful process. In addition to Prof. YILDIRIM and Dr. SAĞLAM, I also want to extend my gratitude to the members of the rest examining committee in the final defence: Prof. Dr. Aslı Özlem TARAKCIOĞLU, Assoc. Prof. Dr. Mustafa CANER, and Assoc. Prof. Dr. Gökhan ÖZTÜRK, all of whom gave support and suggestion for my work.

I also wish to acknowledge Assist. Prof. Dr. Harun SERPİL, who kindly devoted his time and effort to help me through data collection procedure of the study. In addition, I am truly grateful to my friends, Cemre İŞLER and Esin DÜNDAR, whose assistance gave me energy and patience to continue this work.

A very special thank you goes to Prof. Dr. Sezer ŞENER KOMSUOĞLU for support and insightful comments to improve my academic writing process.

My great gratitude goes to my wife, Zeynep ŞENGÜL, and my daughter, Zeynep Sıla ŞENGÜL, for their love and care. Their belief in me and support gave me strength throughout the research. Finally, it would be an injustice to ignore and forget my family, namely my father Nurettin, my mother Meryem, my mother-in-law Neriman, my sisters, Filiz, Ayşegül, and Tuğba, and my brother-in-law Zafer, for their patience during my unbearable times.

Mehmet ŞENGÜL
Eskişehir 2022

STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES

I hereby truthfully declare that this thesis is an original work prepared by me; that I have behaved in accordance with the scientific ethical principles and rules throughout the stages of preparation, data collection, analysis, and presentation of my work; that I have cited the sources of all the data and information that could be obtained within the scope of this study, and included these sources in the references section; and that this study has been scanned for plagiarism with “scientific plagiarism detection programme” used by Anadolu University, and that “it does not have any plagiarism” whatsoever. I also declare that, if a case contrary to my declaration is detected in my work at any time, I hereby express my consent to all the ethical and legal consequences that are involved.

Mehmet ŞENGÜL

TABLE OF CONTENTS

	<u>Page</u>
COVER PAGE.....	i
FINAL APPROVAL FOR THESIS.....	ii
ABSTRACT.....	iii
ÖZET.....	iv
ACKNOWLEDGEMENT	v
STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES.....	vi
TABLE OF CONTENTS.....	vii
LIST OF TABLES.....	xi
LIST OF FIGURES.....	xiii
LIST OF ABBREVIATIONS.....	xiv

CHAPTER 1

1. INTRODUCTION.....	1
1.1. Background to the Study.....	1
1.2. Significance of the Study.....	5
1.3. Purpose of the Study and the Research Questions.....	6
1.4. Limitations of the Study.....	7
1.5. Organisation of the Chapters.....	8

CHAPTER 2

2. REVIEW OF LITERATURE.....	9
2.1. Turkish Higher Education System.....	9
2.2. EFL in Turkish Higher Education.....	13
2.3. Testing and Assessment in EFL.....	18
2.3.1. Types of testing and assessment in EFL.....	19
2.3.2. Key concepts in testing and assessment in EFL.....	21
2.4. Preparatory Language Schools in Turkish Higher Education.....	24
2.5. EFL Language Testing and Assessment in Preparatory Schools	

	<u>Page</u>
in Türkiye.....	27
2.6. Related Studies.....	30

CHAPTER 3

3. METHODOLOGY.....	33
3.1. The First Phase.....	35
3.1.1. Participants.....	35
3.1.2. Data collection tools.....	38
3.1.2.1. Step 1: Literature review, document analysis and informal meetings.....	39
3.1.2.2. Step 2: Initial draft of the interview form.....	39
3.1.2.3. Step 3: Taking expert views.....	39
3.1.2.4. Step 4: Piloting.....	40
3.1.2.5. Step 5: Finalizing the form.....	40
3.1.3. Data collection procedure.....	40
3.1.4. Data analysis procedure.....	44
3.1.5. Validity and reliability.....	48
3.2. The Second Phase.....	49
3.2.1. Participants.....	50
3.2.2. Data collection procedure.....	51
3.2.2.1. Round 1: Written responses.....	52
3.2.2.2. Round 2: Focus group meeting.....	52
3.2.2.3. Round 3: Final approval of the experts.....	53
3.2.3. Data analysis procedure.....	53
3.2.4. Trustworthiness.....	54
3.3. Ethical Issues.....	54

CHAPTER 4

4. FINDINGS.....	57
4.1. Findings of the First Phase.....	57
4.1.1. Systemic challenges.....	59

	<u>Page</u>
4.1.1.1. <i>National challenges</i>	61
4.1.1.2. <i>Institutional challenges</i>	75
4.1.1.3. <i>Unit-related challenges</i>	90
4.1.2. Stakeholder-related challenges.....	102
4.1.2.1. <i>Challenges related with administrators</i>	104
4.1.2.2. <i>Challenges related with testing unit members</i>	110
4.1.2.3. <i>Challenges related with instructors</i>	120
4.1.2.4. <i>Challenges related with students</i>	127
4.1.3. Testing processes-related challenges.....	136
4.1.3.1. <i>Challenges during preparation</i>	138
4.1.3.2. <i>Challenges during implementation</i>	144
4.1.3.3. <i>Challenges during evaluation</i>	150
4.1.3.4. <i>Challenges during post-evaluation</i>	157
4.1.4. General/other challenges.....	165
4.1.4.1. <i>Challenges related with examination</i>	167
4.1.4.2. <i>Challenges related with language education</i>	174
4.1.4.3. <i>Challenges related with time</i>	176
4.2. Findings of the Second Phase	182
4.2.1. National-level solutions.....	183
4.2.2. Institutional-level solutions.....	190
4.2.3. Unit-level solutions.....	197

CHAPTER 5

5. DISCUSSION.....	208
5.1. Key Finding 1: Legal Ground.....	210
5.2. Key Finding 2: Washback Effect.....	217
5.3. Key Finding 3: Standardisation.....	224
5.4. Key Finding 4: Training/Specialisation.....	229
5.5. Key Finding 5: Emotional Factors.....	233
5.6. Key Finding 6: Technical/Physical/Human Infrastructure.....	239

	<u>Page</u>
5.7. Key Finding 7: Sustainability/Efficiency.....	240
5.8. Key Finding 8: Digitalisation.....	243

CHAPTER 6

6. CONCLUSION.....	246
6.1. Summary of the Study.....	246
6.2. Implications of the Study.....	253
6.2.1. Implications for authority bodies.....	253
6.2.2. Implications for EPP Directors and Testing Unit Members.....	256
6.3. Suggestions for Further Studies.....	257
REFERENCES.....	259
APPENDICES	
CURRICULUM VITAE	

LIST OF TABLES

	<u>Page</u>
Table 2.1. The student numbers in THE.....	10
Table 2.2. The academician numbers in THE.....	10
Table 2.3. The language of instruction in bachelor's degree programmes.....	15
Table 3.1. The participants by university type.....	35
Table 3.2. The number of universities by characteristics.....	36
Table 3.3. The number of universities by region.....	37
Table 4.1. Challenges in testing processes.....	57
Table 4.2. Systemic challenges.....	59
Table 4.3. Challenges at the national level.....	61
Table 4.4. Challenges at the institutional level.....	75
Table 4.5. Challenges at the unit level.....	90
Table 4.6. Stakeholder-related challenges.....	102
Table 4.7. Challenges related with administrators.....	104
Table 4.8. Challenges related with testing unit members.....	110
Table 4.9. Challenges related with instructors.....	120
Table 4.10. Challenges related with students.....	127
Table 4.11. Testing processes-related challenges.....	136
Table 4.12. Challenges during preparation.....	138
Table 4.13. Challenges during implementation.....	144
Table 4.14. Challenges during evaluation.....	150
Table 4.15. Challenges during post-evaluation.....	157

	<u>Page</u>
Table 4.16. General/other challenges.....	165
Table 4.17. Challenges related with examination.....	167
Table 4.18. Challenges related with language education.....	174
Table 4.19. Challenges related with time.....	176
Table 4.20. Main categories of solutions.....	183
Table 4.21. National-level solutions.....	184
Table 4.22. Institutional-level solutions.....	190
Table 4.23. Unit-level solutions.....	198
Table 6.1. Summary of the challenges and solutions with regard to key concepts.....	250

LIST OF FIGURES

	<u>Page</u>
Figure 2.1. The number of HEIs	9
Figure 2.2. Turkish national education and higher education system	12
Figure 2.3. The number of English programmes at the undergraduate level.....	16
Figure 3.1. The phases of the study.....	33
Figure 3.2. The steps of the data analysis	45
Figure 3.3. A visual model for the coding process	47
Figure 4.1. The categories and subcategories of the challenges.....	58
Figure 4.2. Systemic challenges.....	60
Figure 4.3. Stakeholder-related challenges.....	103
Figure 4.4. Testing processes-related challenges.....	137
Figure 4.5. General/other challenges.....	166
Figure 4.6. Main categories of solutions.....	182
Figure 4.7. Key concepts.....	206

LIST OF ABBREVIATIONS

EPP	: English Preparatory Programme
YÖK	: Council of Higher Education
MEB	: Ministry of National Education
YÖKAK	: Turkish Higher Education Quality Council
ÖSYM	: The Student Selection and Placement Centre
TÜBİTAK	: Scientific and Technological Research Council of Türkiye
TEPAV	: The Economic Policy Research Foundation of Türkiye
TU	: Testing Unit
TUM	: Testing Unit Member
HE	: Higher Education
THE	: Turkish Higher Education
HEI	: Higher Education Institution
EFL	: English as a Foreign Language
ELT	: English Language Teaching
YDS	: Foreign Language Proficiency Exam
YÖKDİL	: Higher Education Institutions Foreign Language Exam
KPDS	: Public Personnel Foreign Language Proficiency Exam
ÜDS	: Interuniversity Board Foreign Language Exam
CEFR	: The Common European Framework for Languages
EMI	: English as a Medium of Instruction
TMI	: Turkish as a Medium of Instruction

CHAPTER 1

1. INTRODUCTION

This study primarily focuses on the problems regarding examination practices and reaches a consensus on common solutions accordingly. Although the study has some limitations, it is presupposed that it will present some crucial results.

In this part, the study's background and statement of the problem are outlined. In addition to this, the significance of the study, along with its purpose and research questions, limitations, and organisation, is presented.

1.1. Background to the Study

The inadequacy of language education has been a matter of debate in Türkiye (Bayraktaroğlu, 2014; Demir, 2012; Haznedar, 2010; Ulum & Uzun, 2020), about which some changes and revisions have been made (Irgatoğlu, 2017; Kırkgöz, 2007). Of particular concern has been language testing, which has drawn a great deal of attention (Köksal & Cesur, 2012). In order to learn English as a foreign language (EFL), or to prove their proficiency level, which is demanded by private and state institutions, people spend a lot of money, time, and energy. However, is it enough to learn a foreign language, namely English, by spending solely money, time, and energy?

In the study conducted by the British Council and the Economic Policy Research Foundation of Türkiye (TEPAV) in 2014, a needs analysis in a national context of English language teaching in state schools in Türkiye was presented (British Council, 2014).

The findings of the study showed that language teachers could not find opportunities to display their potential in the classroom, course books are heavily used, teacher-centredness is favoured in class with avoidance of learner styles and differences, and particularly learners do not attempt to internalize what is taught but tend to memorize abstract linguistic structures and vocabulary due to examination choices both in class and in standardized tests. In another study conducted by the British Council in 2015, a situation analysis was investigated regarding English in higher education institutions (British Council, 2015).

In the national and international arenas, foreign language education in Türkiye is found inefficient. According to data from the English Proficiency Index in 2022, where countries are ranked every year considering citizens' test scores on language skills from the EF test, Türkiye was ranked 64 out of 111 countries classified as "very low level." Likewise, out of 27 European countries, it was 27th, namely the lowest European country. The highest level was seen in Central Anatolia, while the lowest one was in the Southeast. This index illustrates that English language education in higher education remains a problematic issue in Türkiye (EF Index, 2022).

As mentioned above, the education system in Türkiye has radically changed in recent years. Adopted in Europe, including Türkiye, the Common European Framework for Languages (CEFR) stipulated by the European Council has influenced and made significant changes in the foreign language teaching system. Thus, in foreign language learning, there has been a need for a new learning-teaching approach aiming at improving communication-based skills and six main levels consist of a description of what a language learner can do for each level. Thus, the levels have been named A1-A2 (basic user), B1-B2 (independent user), and C1-C2 (proficient user) and five skills are comprised of reading and listening as receptive skills, and spoken production, spoken interaction, and writing as productive skills. According to CEFR, language learners are divided into three broad parts and six levels: A1–A2 (Basic User), B1–B2 (Independent User), and C1–C2 (Proficient User) (Tannenbaum & Wylie, 2008). In CEFR, language levels are depicted in detail according to skills. In other words, at each level, for every skill a lot of descriptors are developed; thus, it has been clarified that language levels are described via a great number of descriptors. Therefore, at all levels from A1 to C2, the vocabulary storage, grammatical structures, the sorts of conversations and dialogues, comprehension, and expressing yourself at various levels have been put forward in a detailed manner. Thanks to the standards created within the Framework of European Language Policies, it is envisaged to describe foreign language education in a common definition. Accordingly, knowledge and skills of foreign languages in European countries will be identified and documented with reference to A1-C2 levels, and a language learner, no matter where he goes all around Europe, with the language knowledge of one's own, s/he can express and document the level that s/he has had. In that sense, nearly all European countries have adopted the levels and the language education system at schools has been reshaped in accordance with the CEFR. In the

same vein, all of the certificates have either been changed and new certificates with a reference to A1-C1 levels, or been adapted according to these levels. Furthermore, nearly all of the language teaching in the education systems have been prepared with regard to CEFR.

It is also essential to determine which language competencies are needed; for instance, a question should be asked on whether the staff in charge would use language for writing pursuing international relations, communicating effectively with visitors coming from abroad, or reading the archives of foreign origin. It is then possible to find people with suitable language knowledge and skills in staffing. The experts in the field of testing are to be in charge and serve testing institutions and centres. Moreover, the examinations prepared should be accredited and provide international quality standards. For instance, a person with an aim to go on their education abroad should have a certificate valid all over the world, and many other people all around the world should come and get the certificate in our country.

Considering the fact that students in higher education are able to receive courses delivered in English, follow new opinions and movements in their own field of study, interact with foreign researchers or students and seek jobs abroad, proficiency in English is conceived as a needed skill in Turkish higher education institutions in this century. Therefore, language preparatory schools (LPS) in Türkiye mainly attempt to prepare the newly registered students for the courses they will take when they continue their education in their faculties. To do so, a minimum degree of proficiency is required for the completion of these EPPs. Though the completion degree may be justifiable in some universities, it shows excellent variety from one university to other. Hence, a national policy is necessary to set minimum standards of examination and accountability (Pearson, 2021; Shohamy, 2001).

Students having tertiary education are expected to learn English in many HEIs in Türkiye due to globalism and competitiveness in employment. The medium of instruction in English varies from thirty percent to one hundred percent of courses. Thus, universities are required to provide students with English medium courses by the regulations of Council of Higher Education. Though foreign language education delivered in K-12 schools is not within the context of this study, it would be wise enough to mention that the students, having completed their K-12 education, do not

attain an expected proficiency level once they begin tertiary level even though K-12 students start to learn English in the 2nd grade and receive EFL for nearly eleven years.

The students who are required to receive one-year preparatory language education may take examinations to complete the programme, be exempt from language courses and continue to faculty education, or be placed at an appropriate level. Due to the lack of shared regulation and standardisation in language testing, there is a great variety among universities in terms of language level. Therefore, EPPs in some higher education institutions may be pretty challenging; on the other hand, the ones in other HEIs target lower levels for the completion of these programmes. Likewise, language skills and necessities for completion also vary amongst universities and so does the examination. However, due to such inconsistencies, comparability, as well as accountability, remains a great concern for full-time or part-time exchange students. Furthermore, this accountability problem may lead to the recognition and credibility of HEIs in Türkiye and the world.

English language preparatory education is, in general, delivered as an intensive language education at the beginning year of tertiary education. This education has no bonds with the student's registered faculty. The students taking this intensive language education are offered 16-24 hours a week, which is determined according to language level. In addition to course contents, course books, approach(es), methods and techniques, there is great variation in the selection of examination. Generally, an exemption test is administered to allow students with sufficient language level to be exempt from one-year intensive language education and to begin directly to take faculty courses. Following the exemption examination, there are two applications. While in some universities, a placement test is administered to decide and place the remaining students who have failed the exemption test to the appropriate language level, other universities do not administer an additional test and determine the students' language levels according to the score taken in the exemption examination. Detailed information is going to be provided in the next section.

To conclude, though there are some studies on EPPs, they are very limited and problems persist. A policy about accreditation and standardisation is crucial at a national level since the gap between universities is rather extensive. This study aims to point out the problems regarding examination practices and reach a consensus on common solutions accordingly.

1.2. Significance of the Study

This study has several significances in order to fill the gap in language assessment practices. Annual meetings of EPPs are organised with the participation of school directors, where they discuss the challenges encountered at EPPs. Several reports reveal that there are severe challenges in language assessment practices which can be seen to remain from one meeting to another. To date, no study has been conducted about the challenges and solutions in language testing processes and procedures at EPPs in Turkish universities in a comprehensive way. In this regard, this study aims to identify the obstacles encountered in language practices. Another similar significance related to the identification of the challenges, this study was conducted with the inclusion of 50 universities, which is one of the strengths of the study. In addition, some criteria were determined to include different university types, such as characteristics of the university, type of the university, and region factor. This goal-oriented selection of the universities was considered to enhance the probability of inclusion of more challenges, not to leave a significant challenge outside. In addition to selecting different university types, both school directors and TUMs were included in the study, to ensure a different angle.

Besides, the literature review showed that there was no interview form or questionnaire available focusing on instructors' or TUMs' views on testing challenges, representing the status of EPPs in Türkiye. Therefore, the researcher developed an original semi-structured interview form to be used to determine the challenges. The findings gained through the employment of this form during the interviews were aimed at providing a firm basis for developing solutions. Furthermore, almost all the interviews were conducted face-to-face before the pandemic to comfort the interviewees so that they could explain the challenges as much in detail as possible. Otherwise, the participants could have avoided elaborating on some challenges since TUMs could be afraid of any potential risk. In this way, the interviews lasted longer than expected with the help of a clear explanation of the importance and ethical issues of the study.

The solution development will be conducted using the Delphi method. Taking into account the diversity of universities, EPPs, and assessment systems in place, this study is expected to be a significant step towards holistically identifying the common problems and addressing them in a standardized, well-informed, and professional manner.

1.3. Purpose of the Study and the Research Questions

Turkish preparatory schools do not implement a standard procedure to test their students. For example, preparatory students complain that there are inconsistencies in the exams and preparatory instructors complain that they do not have the requisite training in testing (Aydın, 2017; British Council, 2015). There are different perceptions about foreign language proficiency in higher education institutions. To illustrate, although all institutions accept that the content of the proficiency exam should cover four skills, including reading, listening, writing, and speaking, in many institutions, speaking skills cannot be measured for practical reasons. The number of members working in EPPs and trained in assessment is negligible (Aydın, 2017). The findings obtained from the leading state universities reveal differences and problems in the preparation, implementation, and evaluation stages of foreign language proficiency. Based on other relevant reports, it seems safe to conclude that these problems are experienced on a larger scale in Türkiye (Aydın, 2017; British Council, 2014; Kırkgöz, 2017; Unal & Ilhan, 2017).

Further, lending support to Aydın's findings, Aygün (2017) reports that an essential source of student demotivation stems from a high number of tests and a perceived mismatch between what they learn in class and what they are asked in the exams. Such lack of standardisation and confusion in testing in EPPs have recently been pointed out by the President of Türkiye as follows: "we are aiming to attain a certain standard in these exams. This can be done by YÖK for the foreign language proficiency and exemption exams" (Presidential Address, personal communication, 2018). The language proficiency tests administered at the entry and exit levels of the universities are inconsistent and incomparable, which poses a great challenge regarding reaching reliable conclusions about undergraduate students' language proficiency (Güven & Gough, personal communication, 2018). Keser (2018) found a great variation among universities in terms of exit testing as she asserts that EPPs in Türkiye need to cover all four skills in their proficiency tests, especially emphasizing productive skills. As such, the aim of this study is to develop a valid and reliable interview form so as to obtain the opinions of stakeholders and to employ it as a basis for the solution of these aforementioned problems, including inadequate standardisation, lack of consensus on applying valid and reliable exams, and over-testing.

The research studies carried out at preparatory classes in Türkiye have concentrated more on the instruction, students' perceptions, attributions and motivations, success factors, language anxiety, technology adaptation, and evaluation of the programme contents (Balcı & Ügüten, 2018). The review of research literature showed that there was no study explicitly investigating the challenges and solutions in testing processes and procedures at preparatory language schools in Turkish universities in a comprehensive way. Not only does this research study investigate how testing practices are conducted at preparatory classes; but it also attempts to find out the problematic issues regarding evaluation at preparatory classes to be able to suggest solutions by establishing the involvement of different stakeholders.

It is known that there are some differences in terms of terminology between testing and assessment, especially in application. However, the two concepts, namely testing and assessment are used interchangeably in this study. The justification can be the attempt to focus on the practices conducted by TUMs in the TUs. Some testing types will be explained in the literature review section.

This study aims to provide an in-depth analysis of testing systems at preparatory classes in Türkiye by finding out the challenges of the process and producing common solutions to these challenges accordingly.

In this study, the following research questions were addressed:

1. What are the views of language instructors and administrators on the challenges of testing processes at preparatory language schools?
2. What are the suggestions to effectively implement testing processes at preparatory language schools?

1.4. Limitations of the Study

This study has two major limitations, one of which dwells on the population of the study. The study includes language instructors and administrators in identifying the problem and language experts in the solution phase. Other stakeholders are not included as it is thought that this study could be a step for other researchers to extend the population of the study. The other limitation is that 10% of the total universities are included in the research. The universities are classified according to various criteria in order to increase the visibility of different university types.

1.5. Organisation of the Chapters

The present research report is organized into six chapters. Chapter 1 presents the background of the study, statement of the problem, the significance of the study, purpose and research questions, limitations, and chapter organisation.

Chapter 2 provides detailed information on the related literature as well as the studies covering language testing and assessment at EPPs.

Chapter 3 covers the research methodology of the study. It explains who the participants are, what kind of data collection tools are applied, what kind of procedure is followed for the data collection, and what types of analyses are applied.

Chapter 4 presents the findings of the study.

In Chapter 5, the findings are discussed and supported in line with the related literature.

Chapter 6 summarizes the present study, draws conclusions and implications for governing bodies, university/school leaders, and Testing Unit Members (TUMs), and lastly provides suggestions for further research.

CHAPTER 2

2. LITERATURE REVIEW

In this chapter, first, the general overview of the Turkish Higher Education (THE) system will be presented. Second, the place of EFL in THE will be reviewed. Third, EPPs in THE will be introduced, which will be followed by language testing and assessment. Fourth, language testing in preparatory schools in THE will be outlined. Finally, the related literature will be reviewed and discussed.

2.1. Turkish Higher Education System

THE system is legally based on Higher Education Law No. 2547, entering into force on 6 November 1981. THE is planned, coordinated, governed, and supervised by an autonomous public body, namely Council of Higher Education (YÖK), within the provisions set forth in the Law on Higher Education and in the Constitution of the Republic of Türkiye.

In THE, having frequent developments and reforms, 1992 and 2006 are considered two essential breaking points, especially in terms of the expansion of the system (Günay & Günay, 2017). In 1992, the number of universities increased from 29 to 53 with the establishment of 24 new universities. In 2006, the number of universities reached more than 150 with the project “one university for every city.” As of 2021, there are 207 universities in total as shown in Figure 2.1. State and foundation higher education institutions (HEIs) are founded by and subjected to the regulations enacted in accordance with Higher Education Law (YÖK, 2021)

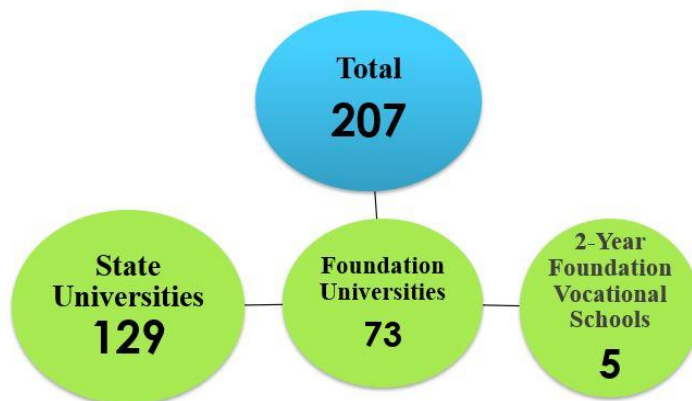


Figure 2.1. The number of HEIs

HEIs comprise two-year vocational schools offering associate's programmes, faculties offering bachelor's programmes, four-year professional schools offering bachelor's programmes and graduate schools/institutes offering master's and doctoral programmes. In addition, Turkish Higher education includes post-secondary higher education programmes, consisting of such as undergraduate-level studies such as short-cycle (associate's), first-cycle (bachelor's) degrees, and graduate-level studies such as second-cycle (master's) and third-cycle (doctoral) degrees (Erdoğan, 2010).

In 207 higher education institutions, more than 8.2 million students receive higher education, 4.04 million of whom are women while 4.2 million are men. In detail, 3.2 million students are enrolled in associate's, about 4.6 million in undergraduate, around 360 thousand in master's, and approximately 11 thousand students in doctorate degrees. Table 2.1 shows the numerical distribution of the students in THE (<https://istatistik.yok.gov.tr/>).

Table 2.1. *The student numbers in THE*

Gender/Degree	Doctorate	Master's	Bachelor's	Associate's
Female	53,361	174,311	2,214,463	1,682,677
Male	56,179	183,960	2,364,584	1,567,424
<u>Sub-total</u>	109,540	358,271	4,579,047	3,250,101
Total	8,296,959			

Considering the number of academicians illustrated in Table 2.2, there are more than 181 thousand academicians in THE. By title, there are 32 thousand professors, 19 thousand associate professors, 41 thousand assistant professors, 38 thousand instructors, and 98 thousand research assistants. One hundred fifty-two thousand academicians work at state universities while 29 thousand work at foundation universities (<https://istatistik.yok.gov.tr/>).

Table 2.2. *The academician numbers in THE*

Gender/Position	Prof.	Assoc. Prof.	Assist. Prof.	Res. Assist.	Lect.
Female	21,922	8,654	19,654	25,436	19,210
Male	11,052	12,694	22,943	23,007	18,358
<u>Sub-total</u>	32,974	21,348	42,597	48,443	37,568
Total	182,930				

The structure of the first and second cycles is separate except for one-tier programmes such as dentistry, pharmacy, medicine, and veterinary programmes, the duration of which is five years but for medicine which lasts six years. The level of qualifications in these one-tier systems is equivalent to that of the second cycle, including the first cycle.

Admission of national students to the short and first cycle degree programmes is centralised, based on a nationwide one/two-stage examination(s) conducted by an autonomous public body, namely ÖSYM. Candidates are placed into higher education institutions based on their composite scores, including their high school grade point averages.

With Türkiye's involvement in the Bologna Process in 2001, THE system has gained a different dimension. In addition to mobility, dimensions such as quality, recognition, and qualifications in higher education have gained importance in the context of policy in Türkiye and worldwide (Erdem, 2013). Regarding mobility, ECTS became compulsory for all HEIs in 2011 with the law numbered 6111. Seminars, workshops, and conferences within Bologna Projects were conducted in different parts of Türkiye as the stakeholders at Turkish universities were introduced and trained on the European Credit Transfer System (ECTS) and Diploma Supplement (DS) labels. As a result, 31 Turkish HEIs were entitled to receive the ECTS Label while 57 HEIs earned the DS Label. Thus, credit-related problems in THE system have relatively diminished, which has fostered mobility and internationalisation of the system (Gündoğdu et al., 2016). Following Commission for Academic Evaluation and Quality in HE (YÖDEK), the establishment of HE Quality Council in 2015 increased the awareness of quality assurance in THE. The universities have been visited and evaluated institutionally by this Council that has also contributed to disseminating good practices and reconstructing the weak points. In the membership process into European Union, THE system has determined the European Higher Education Area (EHEA) as a reference point in terms of organisation and performance evaluation which is not only a legal obligation but also an essential point for Türkiye (Toprak et al., 2019a). With the rigorous efforts of the Vocational Qualifications Authority (MYK) and YÖK on higher education qualifications, the Turkish education system, including HE was referenced and approved in all aspects by the European Union in 2017. However, credits, qualifications, quality assurance, recognition, and internationalisation issues in higher

education are not static but constantly changing (Toprak et al., 2019b). Figure 2.2 illustrates Turkish national education and particularly the higher education system.

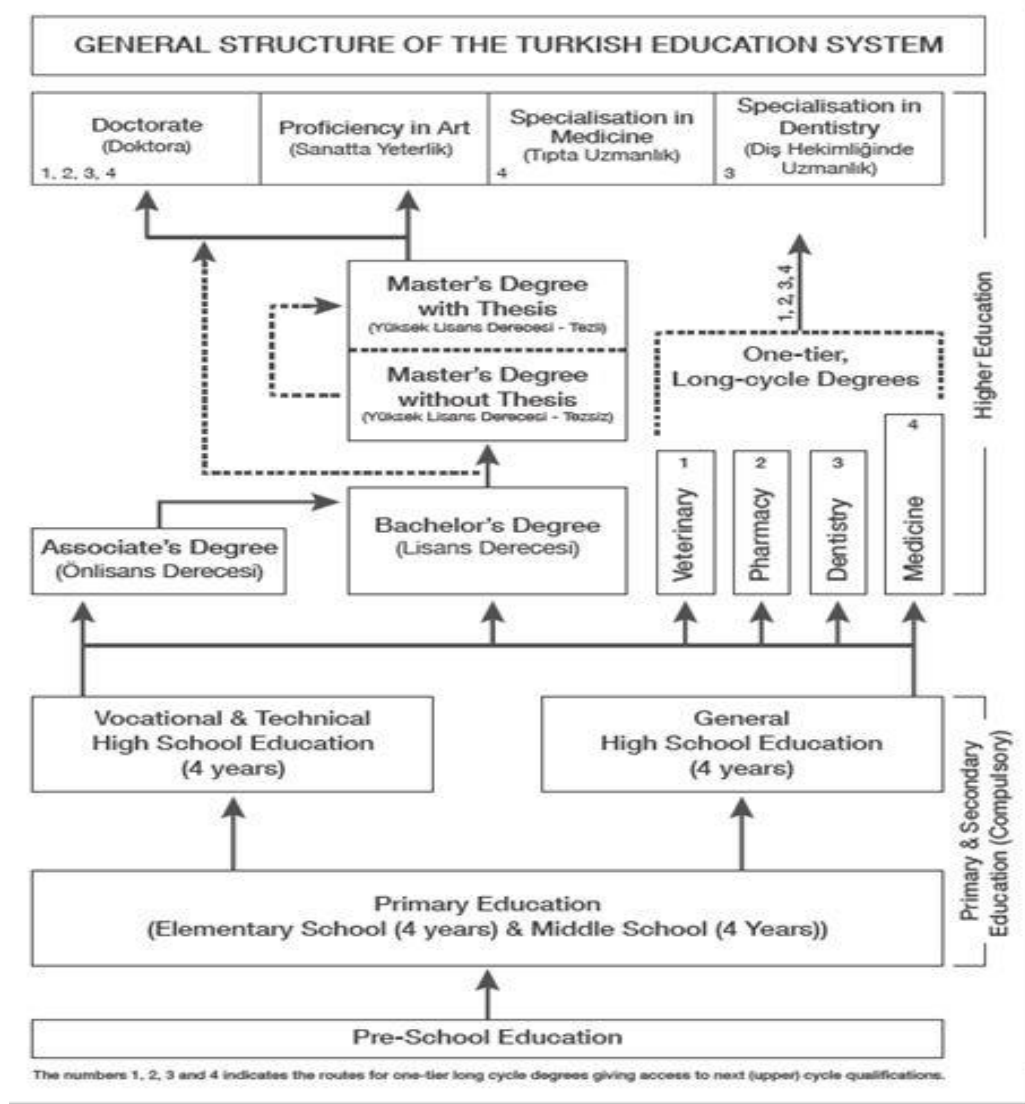


Figure 2.2. Turkish national education and higher education system

With the motto of the New YÖK during 2014-2021, a number of essential projects and studies were carried out in higher education. In 2017, the universities were assigned new missions and visions through the titles of research-oriented universities and regional development-oriented universities. As a result, 23 research-oriented universities and 15 regional development-oriented universities were determined by YÖK. With this model, it is planned to increase universities' efficiency in a research capacity, research quality, and interaction and cooperation (Saraç, 2019). Another major

step in THE, the project, 100/2000 Doctoral Scholarships, was initiated by YÖK, and around 5,000 students have graduated in priority areas within the scope of this project. Thus, qualified doctorates are trained not only for the academy but also for the industry and business sectors in the fields that Türkiye is in need of (Saraç, 2020). With the Internationalisation Strategy Report initiated by YÖK in 2018, strategic targets were determined to transform THE into a centre of attraction within HE. With this report along with a transferable credit system and other initiatives, the number of international students has increased to 240 thousand from 181 countries, which shows how vital internationalisation is in THE system.

One of the significant projects is about the digitalisation of higher education, which has contributed to the smooth transition to digital learning during COVID-19. In the study conducted by Şenay, Şengül and Seggie (2020), despite the efforts by authority institutions, it was found that there are few studies published in quality journals around the world and they suggest that academicians should be fostered to publish their works in high-quality journals. One of the most important aspects of higher education is that its graduates pass into business life and are in national and international competition. Therefore, one of the crucial contemporary problems for graduates is to find a job and have appropriate qualifications for the existing professions, jobs in the sector (McGunagle & Zizka, 2020). In particular, the graduates are suggested to acquire language and communication skills among the 21st-century skills (Ekoç, 2021a). In fact, EFL has a great place in higher education levels, including the undergraduate level. In this regard, the place of EFL in THE system will be discussed in the next part.

2.2. EFL in Turkish Higher Education

Internationalisation has increased the role of English in higher education around the world (Altbach & Knight, 2007). The political issues in neighbouring countries such as Iraq, Afghanistan, and Syria, economy, and domestic politics directly or indirectly have had an impact on foreign language policy, more particularly English. As a result, English is learned, taught as a foreign language, and employed as a medium of instruction in Turkish higher education, making Türkiye an ‘Expanding Circle country’ in Kachru (1999)’s term. The report of ÖSYM indicated that 44 out of 45 international

collaborations were carried out between Turkish HEIs and American/British HEIs in 2013 (Arik, 2020).

The procedures and principles of English language teaching and learning in higher education are laid in the “Regulations on Foreign Language Teaching and Learning in Higher Education Institutions” dated 1984 by YÖK. According to the Regulation on the Principles of Foreign Language Teaching in Higher Education Institutions, the purpose of foreign language teaching is to teach the basic rules of a foreign language so that students can develop vocabulary, understand what they read and hear in a given foreign language, and be able to express themselves in speech or written form (YÖK, 2016). Within this regulation, various changes have been made over the years in order to meet contemporary needs and to provide students with a quality foreign language education at a higher education level. With the regulation numbered 29662 dated March 23, 2016, the latest version is in force. According to this framework regulation, the purpose of foreign language teaching within THE system is to teach students the basic rules of the foreign language, to develop their foreign language vocabulary, to enable them to understand what they read and hear in a foreign language, and to express themselves orally or in writing; and to help graduates of associate’s, undergraduate and graduate degree programmes to acquire foreign language proficiency in their fields. Therefore, EFL is considered to have great importance and place in THE system from different points (Yaman, 2018).

Universities can be divided into three groups in terms of the language of instruction: 1) fully English as a Medium of Instruction (EMI) oriented, where all courses are given in 100% English in all departments, 2) partially EMI-oriented in which the courses are given in 100% or 30% English (or another foreign language) on the basis of a programme. 3) entirely Turkish as a Medium of Instruction (TMI) oriented, where all of all courses are given in the mother tongue; in other words, only Turkish is the language of instruction. According to the above-mentioned regulation, students have to take a one-year English preparatory education before their department education in fully EMI-oriented or partially EMI-oriented universities unless they submit a valid, recognised minimum score. In other group universities, students enrolled in programmes with 30% English as the language of instruction or TMI-oriented ones are able to receive preparatory education upon their request (Öner & Mede, 2015). Research has shown that a number of HEIs in Europe and around the world, have

increasingly adopted English-medium instruction (EMI) in teaching, especially in social sciences and engineering with the impact of internationalisation (Dearden, 2015; Wächter & Maiworm, 2014). Likewise, the number of universities having programmes in which students receive their courses only in English is increasing as they consider studying at such programmes as prestigious for their future careers (Erkuş, Tolga & Sever, 2020; Karakaş, 2015; Kırkgöz, 2014). In other words, the trend in the medium of instruction in higher education has been changing from TMI to EMI or a combination of both (Erarslan, 2019). In the study conducted by (Arik & Arik, 2014), English was found to be the medium of instruction in about 20% of all programmes, at least partially as shown in Table 2.3. Besides, their study showed that it was employed in around 50% of the programmes, especially in engineering and English-related programmes. In addition, they found that most of the newly established foundation universities were English-medium.

Table 2.3. *The language of instruction in bachelor's degree programmes*

Language of instruction	Number of programmes	%
Turkish only	3.593	78,55
English only	723	15,80
Turkish and English	123	2,68
Other	135	2,95
Total	4.574	100

Apart from the medium of instruction, students are supported with English education as a mainstream, a common course in short and first-cycle programmes within THE regardless of their field of study or departments. In the study conducted by Erdem and Atar (2018), it was revealed that the courses English 1 and 2, which are given as common courses, are taught without considering the needs of the students and the socio-economic conditions of the country and the relevant university. They suggest accordingly that these courses should be organized in a more local, need-oriented manner. In addition, vocational foreign language courses, namely English, are given in the programmes where the language of instruction is completely Turkish. Vocational English courses aim to teach and help students learn technical English peculiar to a particular profession that requires specialisation in need for their future career (Gökçe & Batman, 2015). While the content of the vocational English course is directly related to

the needs of the students in their future professions, the content of the general English course may be far from meeting this need (Demiray-Akbulut, 2016). In the study conducted by Özer (2019), acknowledging the indispensable significance of English in higher education, it was indicated that students' perceptions of English and vocational English vary considerably. In this sense, Figure 2.3 illustrates the importance of English at the undergraduate level in Türkiye, with the highest number of programmes (n: 545) delivered in English.

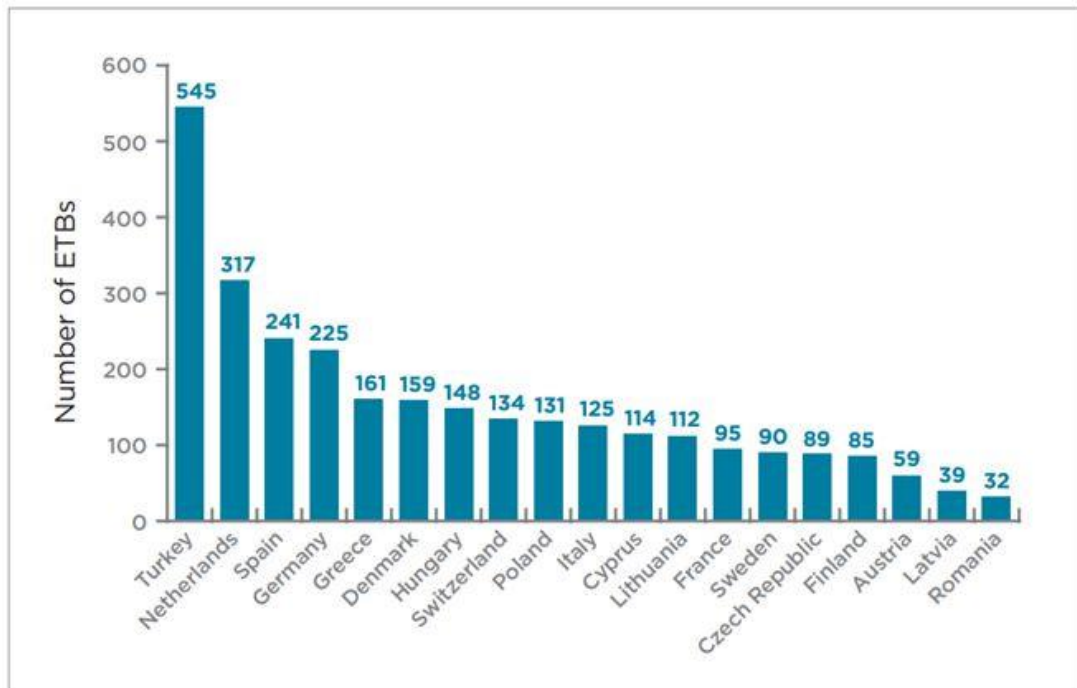


Figure 2.3. The number of English programmes at the undergraduate level (Source: StudyPortals)

The importance of English does not end in associate's and bachelor's degrees as candidates are requested to prove their minimum proficiency levels in order to apply to graduate studies (Doğan et al., 2016). The literature on the importance of foreign language in graduate studies emphasizes the necessity of a foreign language (Bülbül, 2003; Demirtaşlı-Çıktıkçı, 2002; Doğan et al., 2016; Kayahan-Karakul & Karakütük, 2014; Oğuz & Karakaya, 2009). Likewise, in the study conducted by Koğar and Sayın (2014), it was revealed that foreign language score has an important role in postgraduate placements. Within the Regulations for Graduate Studies, while there is no requirement for a score from a foreign language, candidates who would like to apply to doctorate programmes have to prove their proficiency level as the minimum score is 55.

Meanwhile, the universities are given the freedom to set a minimum score for the applications to master's level and to increase the minimum score for doctorate programmes (<https://www.mevzuat.gov.tr>).

The importance of English is not limited to higher education students. Considering a number of journals with high impact factors are published in English, proficiency in English has become a must for scholars worldwide (Bocanegra-Valle, 2014; Selvi, 2021). Therefore, English has become the primary language in international research (Lillis & Curry, 2010) with more than 5.5 million scholars and 2,000 publishers all over the world (Hyland, 2016). Scholars are expected to learn English in order to review literature, conduct their studies, publish and disseminate their studies in English, and pursue their academic careers (Ferguson, Pérez-Llantada & Plo, 2011; Flowerdew, 1999). However, a great majority of scholars are found to experience difficulty conforming to native written English norms, resulting in significant challenges in writing manuscripts as satisfactory as expected by international journal editors and reviewers (Canagarajah, 2005; Flowerdew, 2008). Due to the low level of foreign language among Turkish academics and other factors, the number of predatory journals in Türkiye by Beall increased from seven in 2010 to 41 in 2016 (Akça & Akbulut, 2018). In order to improve the publication quality of Turkish academics, it was decided by YÖK in 2019 that scientific works published in predatory journals cannot be used in academic promotions (Yükseköğretim 2019). In the study conducted by (Mutlu, 2020), it was found that the works of some Turkish academics were rejected due to problems with foreign languages and thus they were able to tend to publish their studies in predatory journals.

Generally speaking, two powerful frameworks are implemented in EFL in THE: CEFR and GSE. However, foreign language education, its criteria, and exam contents are widely designed according to CEFR in THE system. CEFR- based language teaching programmes in Türkiye are increasing in number (Kır & Sülü, 2014). According to the CEFR (CoE, 2020), 80-350 hours of education for A level, 350-680 for B level, and 800-1200 for C level are required. In the study of Çetintaş (2010), it is stated that students take about 650 hours of lessons when they graduate from secondary school. In this regard, graduates from secondary schools are expected to reach B1 level, however, a great majority of studies (Alagozlu, 2012; Albayrak & Şener, 2021; British

Council, 2015; Ekoç, 2021b) have revealed that students starting higher education have weak EFL knowledge and skills.

The report, the 2019 Education First English Proficiency Index illustrates that Türkiye is 79th out of 100 countries with a very low proficiency level (46.81) in the world ranking. It is claimed in the report that the schools in Türkiye spend too much time on grammar and translation in English rather than communication skills (EF Index, 2019). However, In the literature review study conducted by Suna and Durmuşçelebi (2013), they found that there is not only one reason behind the difficulties encountered in EFL in THE. Among the possible issues are the quality of preparatory schools, low student motivation, and a mismatch between students' perceptions and their actual English proficiency level (Arik & Arik, 2018). It has been emphasized in the studies and symposiums in the field of foreign language teaching in Türkiye that problems in language education in Türkiye stem from the problems transferred from primary and secondary education to higher education. Thus, it seems that the English proficiency of the students who are newly enrolled in preparatory language schools at universities is already weak (Ülker, 2017). A great majority of scholars (Akpınar & Cakildere, 2013; Alderson & Wall, 1993; Allen, 2016; Bachman & Palmer, 1996; Cheng & Curtis, 2004; Green, 2007; Külekçi, 2016; Pitoyo, Sumardi & Asib, 2020; Wei, 2017) acknowledge the influence of testing on teaching and learning of English. As examinations are mainly used as means of control and selection of students in Türkiye, they are considered to influence the teaching and learning of English at preparatory language schools.

Considering the impact of tests on language education at EPPs, it would be more appropriate to discuss language testing before examining preparatory schools.

2.3. Testing and Assessment in EFL

The terms ‘measurement’, ‘test’, ‘evaluation’, and ‘assessment’ are commonly employed terms, sometimes interchangeably, in the fields of education, testing, and especially language testing. In measurement, numbers are assigned to qualities or characteristics in harmony with a standard method and the scores are interpreted according to a framework of reference (Genesee, Upshur & Richards, 1996). Testing is perceived as a technical field which is often the measurement aspect of the field such as means, percentiles, standard deviations, and related statistics putting people off (Bachman, 1990; Brown, 2001; Kunnan, 2004). In evaluation, the purpose is to interpret

data gained from measurement to make a judgement on test takers (Bachman, 1990). According to Bachman, Palmer and Palmer (2010), testing, evaluation, and assessment are commonly employed referring to collect information. Although these terms are located on the same paradigm, assessment and evaluation are claimed to be slightly different in practice as assessment can be considered a never-ending process of performance measures (Brantley, 2007) and evaluation as a broader concept involving obtaining and making a judgement on the efficiency of education (Brindley, 2003; Fulcher, 2013). These technical issues leave a lack of confidence that researchers may avoid engaging in testing.

Language testing is considered to have a long history of development (Weir, 2005a) with various classifications such as intuitive, scientific, and communicative by Madsen (1983); discrete point, integrative and communicative by Shohamy (1997); and prescientific, psychometric-structuralist, integrative-sociolinguistic, and communicative by Brown (2005). Language testing is claimed to show similarity with testing in harmony with general principles and measurement in general, but the focus is specifically on language in language testing. Fulcher (2013) stresses the importance of testing in language education claiming that a test is administered with a purpose. The reasons for testing language learners can be to supply students with a degree in the continuum of achievement, to foster further learning, to point out the areas in which they have difficulty, to enhance awareness of the target language, to stress personal aims, to obtain promotion in job, and whatsoever (Birjandi, 2001). Therefore, language testing consists of data interpretation of the students' performance on tasks, by means of which their potential language abilities in real contexts are guessed (McNamara, 2000).

Regarding the appropriateness of the assessment, it is important to be aware of different assessment types which will be explained in the following subsection in brief and select the most suitable ones for the EPP.

2.3.1. Types of testing and assessment in EFL

It is well known that all the language tests are not the same. They vary according to their design, methods, and purposes (Brown, 2004; S. M. Carr, 2011; CoE, 2001; Hughes, 2003; Madsen, 1983; McNamara, 2000). For example, knowledge-based paper-and-pencil tests differ from performance tests according to their method. The first one is mainly administered to check individual items of language such as vocabulary

and grammar or focuses solely on such receptive skills as listening and reading. On the other hand, in performance tests, integration of language skills is included and they are tested with regard to communication (McNamara, 2000).

Achievement tests mainly focus on instructional processes such as the end-of-course examinations, portfolios, or progress observation in the classroom. They are administered to collect proof during a course or in the end to determine whether there has been any progress in defining test-takers' level of mastery (Brantley, 2007) and in which points by supporting learning and teaching within the goals of the course. On the other hand, proficiency tests are administered to measure learners' ability in a particular language (Hughes, 2003) as the purpose of proficiency tests is to conduct measurement of global competence in a given language. Furthermore, in proficiency tests, the earlier teaching and learning process is not referenced as the target lies in future conditions (McNamara, 2000). In these regards, achievement tests pursue objective-based testing goals whereas proficiency tests are administered to assess general mastery.

In discrete-point tests, every linguistic feature is focused on specifically. McNamara describes discrete-point tests as separate, individual points of knowledge; on the other hand, in integrative tests, different language skills such as reading, listening, and speaking are combined (Brown, 2005). "The practice of testing is known as discrete-point testing" (McNamara, 2000). In subjective tests, test-givers' own decisions are applied, whereas specific rubrics are employed in grading objective tests (CoE, 2001). In norm-referenced measurement, a frame of comparison is applied between test-takers to understand the significance of a single score. Information on test takers' achievement is provided in addition to level based on group norms. That is, a score is checked in the light of others, especially in terms of its frequency. In the evaluation of a score in norm-referenced tests, its typicality for the population of other scores is focused. There are some advantages of norm-referenced measurement. One of the benefits could be the definition of levels of performance and recognition of individual performance from others. Another is the well-establishment of aspects and reliability and validity. One disadvantage is argued to be its competitive notion that discourages a number of test-takers (McNamara, 2000).

On the other hand, benchmarks and observable constructs are used for scoring and grading in accordance with course objectives in criterion-referenced tests (Carr, 2005; Davis, 1995; Lum, 2012). Evaluation of individual performance is carried out by

means of verbal descriptions of a satisfactory performance within a certain level. Namely, one's performance is evaluated using descriptors judging the classification of a performance. Unlike norm-referenced tests, quality is regarded in the evaluation of a learner's performance. In this regard, a set of goals are determined for individual test-takers as they are able to succeed at their own rate (McNamara, 2000).

In placement tests, the purpose is to determine learners' level for the suitable group (S. M. Carr, 2011). Learners are grouped into appropriate classes according to their levels (Green & Weir, 2004; Read, 2015). On the other hand, in diagnostic tests, their strengths and weaknesses are identified as testers aim to see what learners know or do not know in a given language (Alderson, 2005; Zhao, 2013). Therefore, the main aim is to identify test-takers' strong and weak points. Diagnostic tests, providing in-depth information related to test-takers' performance, are applied to describe their language profile (Hulstijn, Alderson & Schoonen, 2010).

In summative tests, learners' success is evaluated at the end of the term by reporting the result to an external authority while in formative tests, the process is based (Brindley, 2003; Setiyana, 2016). Test users, especially teachers in a class, administer formative assessment in order to see learners' progress and their level of mastery. In the end, the data gained from such an assessment can be used for planning future courses (Hughes, 2003). Therefore, the summative one is product-oriented, while the latter is process-oriented testing. In productive tests, speaking and writing skills are tested to see learners' performance mainly in the above-mentioned skills.

Considering all the test types mentioned above, the aim of language teaching and learning as well as testing shapes the development, adaptation, and administration of language tests, which necessitates the identification of the purpose in the beginning. In addition, TUMs are expected to regard critical concepts such as validity, reliability and practicality which will be touched on in the following subsection.

2.3.2. Key concepts in testing and assessment in EFL

Developments in language teaching methodology related communicative aspect of language have triggered the emphasis on communicative language testing (Cocchetta, 2018). Based on Hymes's theory of communicative competence, communicative language tests are developed and integrated into language assessment, by means of which test-takers are able to get aware of their progress in communication (Savignon,

2005). According to Harrison (Kostelecká et al., 2015), communicative language tests have three features; assessing language for a purpose, building a bridge of information, and ensuring the representation of an encounter. In communicative language tests, there are two main characteristics, to offer “an extended act of communication” and to assign “social roles” to test-takers.

Within language testing, validity, reliability, test validation, and practicality are key terms that should be taken into account with close attention (McNamara, 2000). Validity shows whether a given test measures what the test is expected to measure. For example, tests of reading are supposed to measure the reading comprehension ability of test-takers while a writing test would test test-takers' composition. With the help of validity, teachers are guided to measure what they teach. However, due to the fact that validity refers to a test's capacity of measuring what it is expected to measure, a number of factors may reduce it.

There are five types of validity: face validity, content validity, construct validity, predictive validity, and concurrent validity. Showing the extent to which a language examination tests the abilities claimed to measure, face validity, including format, arrangement, printing, or typing font, and whatsoever, is the appearance of a test from the perspectives of different observers. To give an example, if a vocabulary test to be administered for the learners enrolled in the Faculty of Tourism consists of terms about pharmacy, it may not look familiar to such students due to its lacking face validity. Thus, it is suggested to show a language test to other teachers, colleagues, or test-deliverers so that they may find out certain ambiguities. Content validity displays the appropriate degree between test content and its level. The content of a test is expected to be in line with what the learners have been taught. Evaluation of the content of the test could be made employing pre-testing and revisions where necessary. Construct validity, claimed as the most significant type as the whole evaluation and measurement create doubts provided that test-providers are unaware of the areas to be tested, refers to what is being tested in an examination. Finally, there is a close relation between predictive and concurrent validity. While predictive validity concerns the predictive force of the tests, the degree to which result(s) predict(s) future outcomes, concurrent validity refers to measures which are already at hand, generally another test. That is, concurrent validity can be ensured only if a test under scrutiny shows a parallel version of a criterion test.

Reliability refers to the consistency of test conclusions and results. There are four methods for ensuring reliability in tests: parallel forms of a test; test and retest; split-half; and internal consistency. The most common way is to administer a similar form of a test while comparing the findings gained from the pilot study. The expected result of reliability is the complete agreement between the two forms. Another way is to employ a test and re-test the same participants after a while. The other one is to deliver a test once but split it into two, regarding it as two different tests. Then, the correlation between the two halves is calculated. Yet, the issue dwells on how to split it into two as there is no guidance. The last one is to consider the internal consistency of a test.

Good tests are supposed to be practical in terms of the feasibility of administration, scoring, and interpretation of the results. Tests can meet reliability and validity, but not practicality. For example, the tests requiring expensive equipment, or certain situations to be conducted, do not provide the ease of administration. Likewise, a subjective test does not have ease of scoring since it may require different raters and lead to fairness doubts. Nonetheless, the function of a test is not suggested to be selected taking into account merely practicality. Namely, a multiple-choice test is not supposed to be chosen for composition rating considering the fact that it is easy to score although we have hired a number of raters. On the contrary, essay tests would be the appropriate way.

Tests have a substantial effect on language education, known as washback effect/backwash effect (Erfiani & Ngadiso, 2018; Fulcher, 2014; Sultana, 2018; Taylor, 2005; Tosuncuoğlu, 2018). It refers to the impact of testing on education and training (McNamara, 2000). The impact driven by testing can emerge in a positive or negative way (Spratt, 2005). It plays a positive impact if language tests support language education, whereas in situations where test content does not comply with course objectives, an adverse effect would take place on language education (McNamara, 2000). To illustrate, the students complain about the variation between what they have studied and what they are compelled to answer during examination on the news in Türkiye after a critical examination. Tests focusing on performance could have a positive influence as language skills are interwoven with language knowledge (Cheng, 2005), whereas multiple-choice exams such as YDS or YÖKDİL administered in Türkiye may have a severe impact on language education. The tests applied are likely to influence not only students' but also teachers' motivation (Guilloteaux & Dörnyei,

2008) as authentic tasks and tests comply with each other. On the contrary, when multiple-choice questions are heavily included, the communication skills of the students are supposed to be restrained since much focus would be given to separate linguistic features (Bellhouse, 2018; Luxia, 2005).

McNamara (2000) found a close relationship between curriculum reforms and the influence of testing on education. He claimed that the testing system is changed by authorities once it is of necessity to generate a new curriculum reform. Meanwhile, he suggested a number of factors such as classroom condition(s), educational habits, motivation, and unforeseen cases which lead to washback influence.

A number of developments have been conducted in language testing in line with the improvements in language teaching. Initially, teachers' intuitive decisions were used to determine the scores or learners' states on the continuum of achievement. Later on, taking into account fundamental qualities and standards, it has been transformed through innovations (Fişne, 2016). Accordingly, there have been a lot of revisions in language testing in Türkiye. It is claimed that testing standards for language proficiency in Türkiye are poor. Learners may get bored, embarrassed, or even irritated as test-takers. In this study, testing is meant for language testing and used as an umbrella term, including assessment and evaluation.

English education is mainly provided to the greatest extent within the preparatory schools at THE. Now, I will turn my attention to these schools, their stakeholders, their missions, their organisational structure, and how they function.

2.4. Preparatory Language Schools in Turkish Higher Education

English preparatory education in Turkish HE has drawn significant attention from various perspectives mainly consisting of the evaluation of the programme (Bayram & Canaran, 2019; Enisa & Uygun, 2014) and curriculum (Aktaş & Gündoğdu, 2020; Mutlu, 2018), motivational factors (Altiner, 2018; Erdoğan & Mede, 2021) and language learning strategies used by preparatory class students (Cesur, 2011; Erdoğan, 2018).

The majority of 207 HEIs have an EPP dedicated to EFL education (British Council, 2015). Although the above-mentioned YÖK directive exists as a framework (YÖK, 2016) regarding the rules to be followed by EPPs, the issues at the implementation level such as passing grades, standards, and use of CEFR are left to the

senates of universities and the administrations of preparatory schools. The students enrolled in the EMI-oriented programmes at HEIs in Türkiye have to study English during the first year of their enrolment at preparatory language schools unless they meet the enrolment conditions of those programmes. Supposed that they have a minimum score from other exams set as equivalent by the Student Selection and Placement Centre (ÖSYM), they do not have to take preparatory education. Once the students who are enrolled in EMI-oriented programmes do not submit a recognised minimum score, they have to take the placement and/or proficiency tests at the beginning of the year. They are required to study for one year (in some universities, three terms) to have a sufficient level of English in intensive programmes. The students unable to meet the requirements of the programmes' final goals within two years are not able to continue their higher education in English-taught programmes.

At English Preparatory Schools, intensive language education is provided for newcomers who have an insufficient language background in English. Learners are expected to be taught English with four basic skills, including listening, reading, speaking, and writing along with grammar and vocabulary considering the main objectives of intensive courses set by (Benseler & Schulz, 1980). According to CEFR, students are aimed to reach B1 level at the end of an academic year (Akpur, 2017). Following the one-year study, the students who obtain a minimum score in the proficiency exam into which other exam results are added for the calculation are allowed to continue to their tertiary education in their faculties.

The general goal of these schools is to help the students to fulfil the language needs which are compulsory for their departmental courses conducted in English (Sarıçoban & Sarıçoban, 2012). Other primary objectives of an EPP can be listed as to teach the students how to read and comprehend so that they can easily follow their courses; to enable them to get the necessary writing skills to be able to take notes and write reports; to be able to listen and speak to follow their lectures; and to be able to ask questions when they go to their departments (Coskun, 2013). Therefore, the students who do not have a sufficient level of English are helped to gain basic language skills so that they can continue their undergraduate education at the university more easily. In other words, these schools both allow students to obtain efficiency in four basic skills of language, namely English, and prepare them for the classes in their departments. In this sense, EPP has a fostering role to ensure a smooth transition to the first year as a next

step and to prepare students for their future departmental courses in various disciplines by helping them develop language skills and strategies effectively. Last but not least, they aim to help students gain insights and create a positive attitude toward learning a foreign language.

As Dearden, Macaro and Akincioglu (2016) suggests, in addition to effective English language learning skills, students have a chance to engage in some additional academic skills such as conducting research, making presentations, listening to lectures and taking notes, and writing academic papers before starting to deal with difficult subjects of coming years. Therefore, EPP is considered to form academic and professional bases students may resort to during their career or further studies.

In preparatory schools, two types of systems are commonly used: modular systems and year-based or term-based systems. The modular system has become one of the most preferred systems recently in order to teach English in not only public but also private schools (Tercan, 2018). Each system provides knowledge and skills towards proficiency in a field (Seçilmiş & Ünlüönen, 2009). Gömleksiz and Erten (2010) remarks that a modular system provides learners with rich content, flexible learning-teaching environment, alternative assessment types, and progress smoothly in compliance with their levels. Students learn at their own pace. M. Öztürk (2015) asserted that the classes are student-centred in the modular systems, which leads students according to their needs and abilities. Various assessment types play a significant role in teaching and learning to recognize learners' achievements in a modular system in which assessment is mainly based on descriptions of the learning outcomes (Coskun, 2013; Ecclestone, 2005). Despite a number of benefits of modular systems, some schools prefer term-based or year-based systems. The entire academic year in the semester-based system does not force the learners to pass between levels (Üstunluoglu et al., 2012).

In English Preparatory Schools, students receive 20-30 hours of English per week for one year based on their level, but this may vary in every university (Akpur, 2017). Their level of instruction changes according to their proficiency in the language. Some give three terms (almost 15 months) but many of them give two terms in English (ten months). If they are optional department students, these unsuccessful students can continue their education in their own departments.

It is a commonly known fact that a majority of the students who exit preparatory schools completing English language education fall behind the language requirements of their departments because of the failures of the preparatory schools in preparing students for their academic life (Coskun, 2013). Besides, British Council's report shows that one-year intensive language education is too short to reach the target language proficiency. According to these studies, the common problem for preparatory education was the mismatch between the expectations of the students and departments from the type of instruction offered in preparatory schools (İnal & Aksoy, 2014; Özkanal & Hakan, 2010). Aydın (2017) also investigated the factors causing demotivation for English teachers and found that ineffective physical conditions were one of the factors that created demotivation. Ulum (2015)'s study illustrated that if basic physical needs in education were not met, language learning quality would decrease. Even though many universities provide English support, many students do not seem to be satisfied with the level of support they receive (Arik & Arik, 2018).

One primary reason for the general failure of these schools lies in the lack of a standardized system among EPPs at different universities. Under current conditions, the level of English proficiency does not play any role in the university entrance system (Öztürk, 2019). Courses in English preparatory schools or units are non-credit, yet students are required to attend classes (British Council, 2015). The biggest deficiencies mentioned by the instructors were determined as low motivation in students, crowded classes, compulsory textbooks, unclear objectives and achievements of the course, and heterogeneity of classes. It is noted that too much course load can affect the performances of the instructors and the content of the courses (Erdem & Atar, 2018). It can be more effective to allocate available resources to infrastructure and pedagogical materials as suggested by Karakaş (2017).

As the purpose of the study is to find out the challenges and possible solutions in language testing at EPPs, the next subsection will include types and practices of language testing and assessment at preparatory schools.

2.5. EFL Language Testing and Assessment in Preparatory Schools in Türkiye

As a necessary component, assessment has an indispensable role in language teaching so that the outcomes of language education can be tested for various purposes. It helps determine whether a topic is understood, and informs the stakeholders of the

progress of a course and learners (Tuzcu-Eken, 2021). In Türkiye, the education system is very examination-oriented. The performances of the students, teachers, and even schools at each ring of the system are evaluated by looking at how well students perform on various exams (Hatipoğlu & Erçetin, 2016). English preparatory schools mostly use institutional tests or TOEFL scores to measure the language proficiency of students (Arik & Arik, 2018). Different measurement tools are used in preparatory schools.

With the proficiency or exemption exams held at the beginning of the academic year, 5% to 25% of the students can go to their departments to start their undergraduate courses without taking preparatory education, and the remaining students are subject to compulsory preparatory education (Akpur, 2017). In the first type of universities, students are required to get a certain score from the exams accepted by ÖSYM such as TOEFL, and YDS, or they are expected to get a certain score from the exams they prepare. Universities usually require at least 70 out of 100. In the second type of universities, compulsory English courses are 30%. These universities, through their senates, can lower the score relative to EMI education. Students are assessed as follows at the end of the eight-week programme: 10% of participation, 20% of achievement exams, and 70 % of the placement exams. There are two achievement tests in each quarter in addition to a writing exam which is a must every week, a listening assessment which is a small pop-up quiz, and a speaking assessment which is an online speaking programme. Students have to take the Exit Exam which includes written and oral assessments at the end of each quarter (Tercan, 2018). It was stated by all the instructors that the assessment and evaluation process included a midterm and a final, and that these exams consisted of multiple-choice tests. Reasons such as the high workload, crowded classrooms, grammar-focused lessons and tests play a crucial role in the determination of multiple-choice tests (Erdem & Atar, 2018).

Bayraktaroğlu (2014) claimed that the evaluation criteria applied in preparatory schools are of a subjective nature and that exemption or qualification criteria should be determined with national or internationally valid measurement tools instead of the exams prepared by the institutions themselves. Similarly, Aydın et al. (2016) state that the preparation of the proficiency exams within the schools causes some problems. It has been found that the measurement tools do not have certain criteria as a common content (Akpur, 2017). Atar, Kır-Cullen and Denkci-Akkaş (2020) believed that the

students found the online assignments insufficient due to the technical issues they had to cope with throughout the academic year. The instructors shared the same issue as well, and they suggested not integrating online assignments into the programme anymore. Ünal, Sarı and Gürol (2017) stated that more technology should be integrated into the curriculum, but it is essential to remember that too many technical problems may frustrate learners and instructors, resulting in negative attitudes. It was revealed that both students and instructors were highly dissatisfied with the assessment and physical conditions of the preparatory programme (Özüdoğru, 2017; Tekin, 2015). The placement scores required by the universities might be too low for students to function successfully in their classes conducted in English (Arik & Arik, 2018). Level progress and student development are all considered to be related to CEFR-exam-material alignment. A probable inconsistency in the assessment and the content covered during the courses may arise and this may be linked to a decrease in students' motivation and willingness to learn (Erarslan, 2019). On the other hand, there is also criticism of the use of CEFR in language testing as Weir (2005b) contends that since the CEFR was not designed specifically for language testing. Öz and Atay (2017)'s study with instructors teaching at the preparatory programme revealed that there is a mismatch between Turkish EFL instructors' in-class assessment literacy and its reflection in practice.

Exams and performance assessments need to be improved regarding extra-curricular activities and online programmes (Bayram & Canaran, 2019). It is a situation in which there are serious differences between levels within the preparatory schools (Akpur, 2017). Some improvements needed to be made, particularly in the assessment system (Özüdoğru, 2017; Tekin, 2015). The mismatch between students' actual level and their placed level in language education is prevented with a valid and reliable placement test, it provides increased student learning and retention (Gonzales, 2011). As also asserted by Cephe and Toprak (2014), exams taking CEFR descriptors into core try to assess what a learner at a certain level can do. this may be solved by aligning level descriptors, namely outcomes, and testing and assessment criteria (Erarslan, 2019). Fulcher (2004) also argues that the CEFR may be of use in language testing as a user-oriented scale which serves as an understandable, practical reporting instrument for stakeholders. Bechger, Kuijper and Maris (2009) suggest that providing a descriptive system of language activities involving different levels of proficiency could be used for

existing tests and examinations. There should be a close relationship and consistency between the teaching and testing processes (Mutlu, 2018).

In the next section, related studies on language assessment conducted in Türkiye will be explained.

2.6. Related Studies

The literature review shows that the studies are not directly related to the problem of the study and that the issues with regard to language testing at EPPs are discussed in the studies on the evaluation of the preparatory programmes as a subdimension.

Şahinel (1997) conducted a research study on English instructors' views on the English language testing conditions in the preparatory classes. The study indicated the techniques employed in the test were not prepared in the order of difficulty and the objectives of the curriculum were not taken into account by test-providers while constructing their tests.

In a study conducted by Serpil (2000), content validation of the exams administered at the preparatory school at Anadolu University was investigated. He found conflicting results caused by the insufficient knowledge of language instructors on testing and their unclear objectives. At the end of the study, Serpil puts forward a number of suggestions in order that testing practices could be improved at EPPs.

Aksan (2001) aimed to examine teachers' perceptions of the relationship between the content of test items and the content of the course book and their teaching. The results of the study indicated that although teachers were satisfied with the content of tests, there were a lot of problematic items in the tests.

In the study of Köksal and Cesur (2012), they aimed to investigate language instructors' and students' perceptions of the exams administered at the compulsory English Course which students receive three times a week at Çanakkale Onsekiz Mart University. The results of the study show that not only students but also teachers favour other test techniques apart from multiple-choice tests that are commonly administered. It is indicated, as well, that differing levels of the students hinder their potential causing them to forget even what they know. In the end, they offer recommendations for the construction and implementation of better tests for students taking this course.

Ersoy and Yapıcıoğlu (2015) conducted a study on the views of students and instructors in an optional English preparatory programme. The findings of the study revealed that both students and instructors viewed the number of exams is too high. They also concluded that the high number of tests reduced the desire of the students. Another point was that the instructors viewed the tests negatively due to constantly preparing exams and evaluating papers, which increased their workload. They also stressed that there are different opinions regarding the difficulty level of the exams as easy or difficult.

Akpur (2017) conducted a study on the evaluation of the preparatory education to find out the problems that students face. He found that despite the fact that the exams were parallel to the course topics, the majority of the participants had a lack of motivation and did not study to learn English. He concluded that the absence of English questions in the university entrance exams negatively affected the success of the preparatory programmes.

In a study carried out by Balcı, Durak-Üğüten and Çolak (2018), the effectiveness of a compulsory English preparatory programme at a state university was investigated. The results of the study illustrated that less emphasis was put on speaking skills as the instructors placed excessive stress on grammar. Another conclusion of the study was an overemphasis on grammar can be resulted mainly from a lack of materials.

Özusu (2018) conducted a study on the exams that were administered at a preparatory school. The findings of the study showed that there was a mismatch among different quizzes, midterms, and final exams. Another finding was that some subjects could not be included in the exam due to the short duration of the second-term quiz. She also found an inadequate number of quizzes assessing non-book activities, a lack of diversity in question types asked in the quiz, and excessive multiple-choice exams. Another problem was the lack of an assessment system to evaluate students' class performance. In addition, they stressed that the level of exams below / above the students, insufficient number of vocabulary questions, and insufficient number of questions assessing the use of language are among the problems concerning the exams. With regard to the writing exams, it was stated that the students used to enter the exam by memorizing the text they were going to write and they had a lack of opportunity to give feedback to the paragraphs of the students. She also criticised that in exams, questions assessing language use were fewer than the questions assessing grammar and

vocabulary due to the fact that the instructors found grammar and vocabulary questions easy to evaluate. The last challenge was that the instructors worried about the objectivity of such assessment techniques.

The study by Keser and Durmuşoğlu-Köse (2019) attempted to identify the exit criteria in English Preparatory Programmes at universities. The results of the study showed that the CEFR descriptors were too broad and vague. They suggested that the students should be provided with more realistic goals.

In another study conducted by Ölmezler-Öztürk (2019), it was aimed to develop and validate the language assessment knowledge scale as an instrument to measure the language assessment knowledge of 542 EFL teachers. She found that the teachers were the most knowledgeable in assessing reading whereas they had the lowest score in assessing listening. Depending on the qualitative data, it was revealed that training as pre-service and in-service was insufficient; thus, teachers were found to be in need of training on assessing skills.

To conclude, the literature review shows that language testing seems to attract the attention of the scholars, yet the studies mainly investigated different issues related to language assessment, but not challenges encountered in EPPs. It is evident that higher education has a key place in career planning and the rest of university students' lives. Regarding the inefficiency of K12 in language education, education received in an EPP can be considered a final stop for learning English before work life. Therefore, EPPs appear to play a crucial role in learners' education in their departments. The students who succeed in language skills during their EPP education are expected to better in adapting to departmental terminology and courses. However, the report conducted by British Council shows the state of English in higher education is not at an expected level. This is supported by EF Index results and other studies. Last, considering the impact of washback on language education, language assessment planning, organisation and units should be supported. That's why this study aims to determine the challenges in assessment practices and common solutions. In the next section, the methodology of the study is explained.

CHAPTER 3

3. METHODOLOGY

The current section presents methodology beginning with the research questions and design of the study. It continues with the participants, the role of the researcher in the study, data collection instruments, and procedures. The chapter ends with detailed information on data analysis and trustworthiness of the study.

The data for this study were obtained during 2019 February-2020 June, and this intensive process lasted sixteen months. The present study consists of two phases each of which is shown in Figure 3.1:

- i. the first phase: interviews with directors and/or vice directors of EPPs, and with testing unit members in separate sessions, in relation to the first research question,
- ii. the second phase: Delphi method with English language experts specialized in language testing, in relation to the second research question.

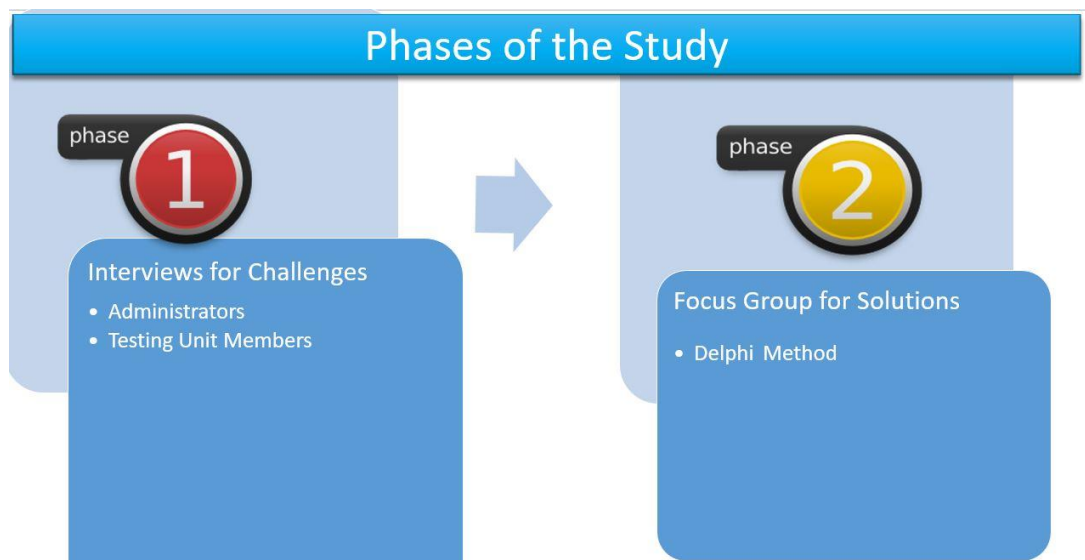


Figure 3.1. *The phases of the study*

Considering the fact that this study is composed of more than one phase and structured with various methods, participants and instruments, methodology is explained, including both phases.

Taking into account the complexity of the issue, this study adopts a qualitative technique in which the collection of data is carried out to make a description of a situation to explore and identify assessment processes at EPPs. As a qualitative data elicitation, participants were asked to answer open-ended questions to reveal their opinions on the challenges encountered in the assessment procedure. In qualitative research, social phenomena are investigated through a number of data collection methods, including interviews, document analysis, and observation generally in their naturally occurring social environments (Yıldırım & Şimşek, 2006). In addition to Denzin and Lincoln (2011), Creswell (2013) presented major characteristics of a qualitative study: a natural setting, researcher as a key instrument, multiple sources of data, inductive data analysis, participants' meanings, holistic account, theoretical lens, interpretive inquiry, and reflexivity of the researcher. Hence, qualitative research data is considered to be more comprehensive providing more details about an issue (Denzin, Lincoln & Giardina, 2006). Patton (2014) also agrees that the researcher is likely to form deep knowledge about the participants and relevant issues in a research study. Besides, qualitative research relies more on factors influencing the results of the study. On the other hand, some issues regarding the weaknesses of a qualitative study are listed, including sample size, generalizability, researcher role, time consumption, and intensive labour (Cheng & Dörnyei, 2007).

The case study was adopted in this study as the design of the study. With this design, an intensive study was aimed at a person, a situation, a group of people or a unit, which could be generalized over several other units. Due to the fact that one important feature of cases is any emphasis on the significance of an analytical frame in the constitution of the study (Thomas, 2011), this design was selected in the study. Case study is explained in an appropriate manner by Simons (2009):

“Case study is an in-depth exploration from multiple perspectives of the complexity and uniqueness of a particular issue or institution in a real-life context (p.21).”

Initially, the research question(s) are provided. Then in each phase, an explanation of the participants, the procedures of data collection, and the data analysis methodology are presented.

3.1. The First Phase

In this section, only the participants, data collection procedure, and data analysis employed in the first phase are presented, respectively.

3.1.1. Participants

Within the context of this research, various participant groups were included. All the participants were explained the purpose and design of the study and a ‘consent form’ was administered to each participant (See Appendix 2). For the sake of anonymity of the participants, they were assigned acronyms such as P1U1T1 (for phase 1, university 1 and testing unit member 1; P1U1D1 (for phase 1, university 1 and director 1), etc. The population of the first phase consisted of two participant groups, one with EPP directors and/or vice directors, and the other with testing unit members at the EPPs. Both groups were interviewed to identify the problems in testing practices in EPPs at tertiary education. The participant groups are explained in detail in Table 3.1.

Table 3.1. *The participants by university type*

University Group	Number of Universities	Number of TUMs	Number of Directors
State	35	95	53
Non-profit foundation	15	40	20
Sub-total		135	73
Total Participants	50		208

As Table 3.1 illustrates, 208 participants from 50 universities (35 state universities and 15 non-profit foundation ones) were included in the study. Of the 208 participants, 135 were TUMs and 73 were school directors.

A number of studies have been carried out in EPPs at Turkish higher education institutions so far, yet there has not been a study categorizing the universities as done above. Çatı, Kethüda and Bilgin (2016) claimed that competition between universities have increased due to globalisation and internationalisation, which results in great differences among the universities in Turkish higher education institutions in terms of academic staff, finance, location, mission, student population, and quality standards, etc. Likewise, this difference exists in their faculties and units, namely in EPPs. Thus, such

a classification was created for this study, to include a variety of universities related to research questions without ignoring some issues peculiar to a university type. For Tekneci (2016), there will be a major difference between research and mission-oriented universities and others. Hence, one category is whether the university is a research-based, mission differentiated, or regional development-oriented higher education institution. Another category is related to the ones having received an external quality assurance considering that if a university has received external quality assurance review, an external agency has reviewed various processes and procedures of all the units involving student participation, lesson plans, administration work, and especially testing documents and procedures (<https://yokak.gov.tr/>). The third criterion is to take the ones providing all courses in 100% English. Since all the students are prepared for their continuing education in their departments in 100% English, these universities are supposed to give more emphasis on foreign languages. Furthermore, the courses provided in EPPs at these universities may deliver English for Specific Purposes or English for Academic Purposes during intensive education. In the fourth group exist the newly established universities established after 2015. (Kılınç et al., 2017) and Doğan (2013) found that the newly established universities appear to have different problems from other universities. The year 2015 is the cut-off point when a number of universities were established together. The last group includes the remaining universities, which do not fall into any category mentioned above. Table 3.2 illustrates the number of universities by their characteristics as explained in detail above.

Table 3.2. *The number of universities by characteristics*

University Group	Number of Universities
Research-based universities/regional development/mission-differentiated universities	5
The ones having received external quality assurance	10
The ones providing all courses in 100% English	4
The newly established universities	7
Other universities	24
Total	50

Table 3.2 shows the number of universities by characteristics. Out of 50 universities included in the study, ten universities have received external quality assurance, seven universities are newly established which appears to have different challenges, five universities are research-based or regional development/mission-differentiated universities selected by YÖK and four universities provide all courses in 100% English, in other words, they deliver EMI-oriented education. The rest of the universities (n=24) are general ones, which means these universities are outside of such classifications. In addition to the characteristics of the universities, they were also distributed in terms of their place of the region. The numbers of the universities by region are illustrated in Table 3.3.

Table 3.3. *The number of universities by region*

University Group	Number of Universities
Aegean	7
Black Sea	5
Mediterranean	7
Marmara	14
East (+South East)	7
Central Anatolia	10
Total	50

As Table 3.3 illustrates, universities were selected from 6 regions. The density of the number of universities in the relevant region was taken into account while determining the selection. Out of 50 universities, 14 were selected from Marmara region, ten from Central Anatolia, seven from Aegean, Mediterranean, and East (+South East) regions each, and five from Black Sea. The regions East and South East were combined due to fewer universities compared to other regions. Stratified sampling was employed in the selection of the universities as they were determined on the above-mentioned specific characteristics. In such a sampling, the sample is guaranteed to consist of specific characteristics the researcher desires included in the population, and the stratum aimed is represented in proportion to the existence of the population.

3.1.2. Data collection tools

In this current study, the instruments for the qualitative data elicitation which were developed by the researcher himself were explained in this section.

To elicit qualitative data, the researcher employed an interview form, including seven open-ended items through which opinions of the participants were gathered on the problems encountered at EPPs. In open-ended interviews, the participants provide their own responses to the questions instead of response options (Klassen et al., 2012). The interview form (See Appendices 3 and 4) was made up of two sections:

- demographic information,
- open-ended questions for the comments and opinions of the participants.

Basically, there are three types of interviews: unstructured, structured, and semi-structured. In unstructured interviews, the interview begins with a more general question enabling the participant to have control in terms of the sequence and the content of the interview. On the other hand, in structured interviews, the content, and procedures of the interviews have been organized beforehand. In the third type, namely semi-structured interviews, the questions are created in advance, yet the interviewer possesses some freedom to make elimination, adjustment, or addition (Johnson & Rowlands, 2012). In this study, a semi-structured interview was employed and the questions were written by the researcher depending on the results of the literature review, document analysis and informal meetings, and focus group discussion. Raworth et al. (2012) noted that semi-structured interviews create a chance for the researcher to respond to the situation under investigation and bring about new ideas on the issue while the participant feels free to express his/her ideas with a certain framework created by the interviewer. Likewise, according to several researchers (Charmaz & Belgrave, 2012; Granot, Brashear & Motta, 2012; Roulston & Choi, 2018), the interviews are not tightly structured in qualitative research as the target is to make participants volunteer to comment in their own terms. Thus, the researcher did his best to make the interviewees tell all the things related to the research topic so as to be able to gather as much data as possible.

3.1.2.1. Step 1: Literature review, document analysis, and informal meetings

Following the determination of the interest construct, a literature review has been done to check whether a relevant interview form has been previously developed. Although similar forms seem to exist in other studies, they are found inappropriate to the context and current situations in Turkish universities. Thus, the researcher has decided to construct a semi-structured interview form which would reflect the common problems encountered in EPPs of Turkish higher education institutions.

Through literature review and document analysis, key points were noted to a list. Document analysis included scanning of websites of EPPs and testing and assessment sections in official guidelines or regulations. Meanwhile, informal meetings and interviews were conducted with some language instructors working at testing offices at universities in Ankara in order to generate a draft programme for interviews. During these meetings, an attempt was made to prepare a draft of issues and cases using the notes extracted through the two processes.

3.1.2.2. Step 2: Initial draft of the interview form

In the next round, according to the categories of ideas and issues encountered in testing practices, items were written to build an item pool. There are various guideline suggestions to write interview items (Artino Jr et al., 2014). An item is expected to be simple and short in addition to assessing a single issue so that target participants are able to comprehend it. The researcher attempted to avoid leading questions since such questions might cause participants to give biased answers (Krosnick, 2018).

When the item pool was prepared, each item was reviewed with the advisor at the end of which the draft of the interview form was shown to some language instructors to receive their suggestions before going through an expert panel. In light of the suggestions, necessary revisions were made.

3.1.2.3. Step 3: Taking expert views

Once a draft of the interview form items pool was prepared, qualified experts conducted reviews of each item in order to ensure that the items were accurate, grammatically correct and did not have item construction problems. They also attempted to make sure no items contain offensive or biased content against a certain

subgroup of participants. In the study, five experts who have been giving testing courses in English Language Teaching Departments in Turkish universities were presented with the draft of the interview form as they were conceived to be acquainted with the topic. They submitted their suggestions in writing form regarding its accuracy, scoring, the length of items, the use of appropriate language, and the place of items.

3.1.2.4. Step 4: *Piloting*

Before administering the interview form to the target population, it is advised to check the questions to a small group (Perneger et al., 2015). The piloting process allows the researcher to get aware of whether confusion appears related to interview items and to receive revision suggestions from the participants in order to review and improve each item where necessary. What is more, piloting may create a chance for the researcher to reach rough conclusions on the distribution of answers, which is helpful to see the extent to which variation fluctuates in the responses so that the researcher could decide to move forward to conducting a large-scale study (Robins & Eisen, 2017). Each item was revised according to the outcomes gained from the piloting stage. The analysis of the pilot study was also paid serious attention since the researcher did not intend to lose valuable data; therefore, this part was not thrown away.

3.1.2.5. Step 5: *Finalizing the form*

At the end of all the steps explained above, necessary revisions were made in the light of suggestions of the experts and participants, as well. Through the last reviews and changes, the interview form was finalised.

3.1.3. Data collection procedure

The participants, namely directors and/or vice directors in one session and testing unit members in the following session, were interviewed through a semi-structured interview form developed through informal meetings with language experts working at Turkish universities and a focus group discussion meeting. A qualitative study is “a type of educational research in which the researcher relies on the views of participants and asks broad, general questions” (Creswell & Creswell, 2005). Berger and Karabenick (2016) stressed that the researcher should pay close attention to whether the questions

are comprehended correctly and whether they are not vague; otherwise, the participants could touch on different but irrelevant aspects of the issue.

During the pilot study, the researcher followed only official process by sending e-mails and official letters to EPPs directorate. However, only one university director replied to the request while that one was negative due to the sensitivity of the research topic. Once the school directors heard that the study was about language testing practices at their school, they would not dare to listen to detailed information, but reject automatically. Therefore, in the main study, the participants were identified through the researcher's friends and professional connections - administrators, professors and ex-colleagues. The second strategy the researcher used to identify and contact potential participants was through participants directing him to others. Some participants were generous in providing names and directing him to three or four of their friends. The researcher only contacted referrals after they were told to expect a phone call or e-mail from the researcher. This strategy eased his job in the sense that they were already expecting a phone call or e-mail about the study. Via these two chains of referrals as vehicles to identify the participants, the researcher was able to refer to a common contact that established some degree of rapport and trust. This gave him at least a chance to explain who he was and what his research was about in detail over the phone or face-to-face before potential participants made a decision about their potential participation in the study (Clark, 2006). These two strategies the researcher used to reach the participants allowed for a purposeful and snowballing sampling (Creswell, 2005). Snowball sampling was particularly useful since some degree of trust was imperative to establish the initial contact. Administrators, ex-colleagues, and participants played a critical role in helping him gain access to the interviewees (Clark, 2006).

Information was gathered by the researcher through site visits to universities and personal interactions with individuals and then inductively generated "meaning from the data collected in the field", which is an interpretation "shaped by the researcher's own experiences and backgrounds" (Creswell, 2005, p.9). Almost all of the interviews were conducted face-to-face during site visits to the related university before the breakout of COVID-19, after which the participants from the universities located in the Black Sea region were decided to be interviewed online or via phone call (to be held in the first week of June). The participants were observed to feel freer to explain the issues in the face-to-face interviews with the researcher.

All the testing unit members and most of the directors are experts in the target language, namely English. However, during the pilot study, it was found that some directors did not graduate from the ELT department. Thus, the interviews were conducted in the participants' mother language; namely Turkish, to prevent any slight language barrier from sharing their comments.

Before the interviews, the participants were asked whether to record the interview. Upon their approval, interviews at 43 universities were recorded by the researcher in order to be transcribed following the interview to ease the data analysis procedure. In the remaining universities, the researcher took notes by handwriting. It was rather challenging for the researcher to keep up with the pace of the interview by not only leading the interview and taking note of the crucial comments, as well.

The participants were personally invited by the researcher and kindly asked whether it was possible for them to participate in this study after being told about the purpose and necessity of the study not only for the quality of the assessment at EPPs and doctoral dissertation of the researcher. In addition, they were assured that their ideas would be confidential during and after the study. All of the participants agreed to take part in the study stating their willingness to participate in such a crucial study and signed the consent form prepared by the researcher.

It was rather tough to lead the study during the first interviews as the researcher faced some major challenges such as time management problems, making false schedules in big cities by looking only at the map, insufficient sleep before the interviews, unable to reach the directors in charge for permission, avoiding disruptive talks and misleading the participants, making to the point clarifications. However, as the researcher gained experience through ongoing interviews, significant measures were taken one by one to lead the interviews more effectively. Below are the measures in relation to the trouble during the interviews:

- Time management problems: Initially, the researcher tried checking his mobile phone, yet it was felt that this distracted the interviewees. Afterward, the researcher put on a wristwatch and attempted to check the time without leaving any sign of it.
- Making false schedules in big cities by looking only at a map: Especially in Istanbul and Izmir, taking the university's venue from their internet address and looking at the map misled the researcher since some universities were moved

but their address was not changed or some universities looked close to each other, but perhaps with a plane or jet. Thus, the researcher started to make programme with a friend from the related university who both know the updated venue and the city conditions such as distance and public transportation.

- Difficulty in reaching the director in charge for permission: The researcher initially tried to get in touch using the phone number of the school displayed on their website. But, in some schools, neither the researcher could find anyone answering the phones nor the secretaries returned a response though they said they would respond soon after checking with the director. They would not take part in the study considering the sensitivity of the topic and the time they and testing unit members would give for the interviews considering their dense programmes. In these universities, the researcher attempted to find a gatekeeper in order to reach the director of the school and provide necessary information on the aim and significance of the study. The researcher took advantage of his friends working at the related university or the key leaders he met during high-level meetings.
- Fatigue of the researcher: The researcher tried going to the destination city one day earlier in order to have a good sleep and breakfast before the interviews.
- Avoiding disruptive talks and misleading the participants: The researcher wrote under the interview questions what is expected from the questions in order to stick with the research purpose. Meanwhile, the participants were free to jump back to answering the earlier questions but they were asked a sub-question once they started to mention off-the-topic issues.

The interviews were conducted outside of the participants' class time or work burden in their offices. Especially having a chance to observe the testing unit rooms during and after the interviews was quite valuable for the researcher to make empathy and sympathy with them.

The interviews with directors and/or vice directors lasted almost from 45 to 60 minutes, and the ones with testing unit members lasted around from 60 to 120 minutes.

It was also tough for the researcher to make arrangements for the interviews since testing unit members were always busy preparing for the upcoming exams in addition to finalising the previous ones, especially at the beginning and end of the academic year.

So, the interviews were carried out at a suitable time to serve the purpose of this study. The interviews with the directors and/or vice directors were conducted for triangulation and comparative purposes; therefore, such interviews were not the main focus of the research.

3.1.4. Data analysis procedure

The data gained from the semi-structured interviews were analysed and interpreted in accordance with the Constant Comparison Method suggesting certain systematic procedures in sets for gathering and analysing the data. The researcher adopted a data-driven approach in order to allow new codes to emerge from the data and categories or themes to emerge from the codes so that a new theory could be generated based on the perceptions and experiences of the participants interviewed.

The researcher used an inductive approach going from particular data to general codes and themes as the final goal was to generate a larger consolidated picture of the challenges. The Constant Comparison Method (Butler-Kisber, 2018; Kolb, 2012; Leech & Onwuegbuzie, 2011) was employed in the whole process of data analysis based on the conceptual framework and the research questions so that general categories were inferred from the specific codes through comparing the emergent codes with the existing codes and categories rather than following predefined categories (Klassen et al., 2012; Kolb, 2012). Employing this method, the researcher is required to use multiple steps in a cyclical manner by segmenting the data into small pieces, as one unit of data is compared constantly with another in order to determine the recurring codes and create categories (Leong, Joseph & Boulay, 2010). In vivo and descriptive coding techniques were primarily employed to determine the codes (Miles, Huberman & Saldaña, 2018). Through in vivo coding, the researcher assigned labels to the units of the data employing actual words or phrases of the participants. In addition, descriptive coding was used by labelling the units of the data and summarizing the main points cited by the participants.

The researcher employed a simultaneous process of analysing while also collecting data or transcribing other audio recordings. The process was iterative as the researcher cycled back and forth between data collection and analysis to fill gaps in their narratives. As this process was labour-intensive, taking time, the researcher was keeping duplicate copies of all forms of data by uploading the data into a folder on drive

and updating the folder frequently. The sample transcribed data of an interview can be found in Appendix 7. The interviewers' questions, prompts, and comments were not coded since the researcher prioritized the participants' data in the analysis of the interviews since he was studying their perceptions, not his. Thus, his interpretations of their narratives via coding were his contribution to the meaning-making enterprise (Saldana, 2011). For data analysis, Miles and Huberman (1994)'s, Creswell (2013)'s and Miles et al. (2018)'s recommendations were taken into consideration. The steps of analysis are illustrated in Figure 3.2.

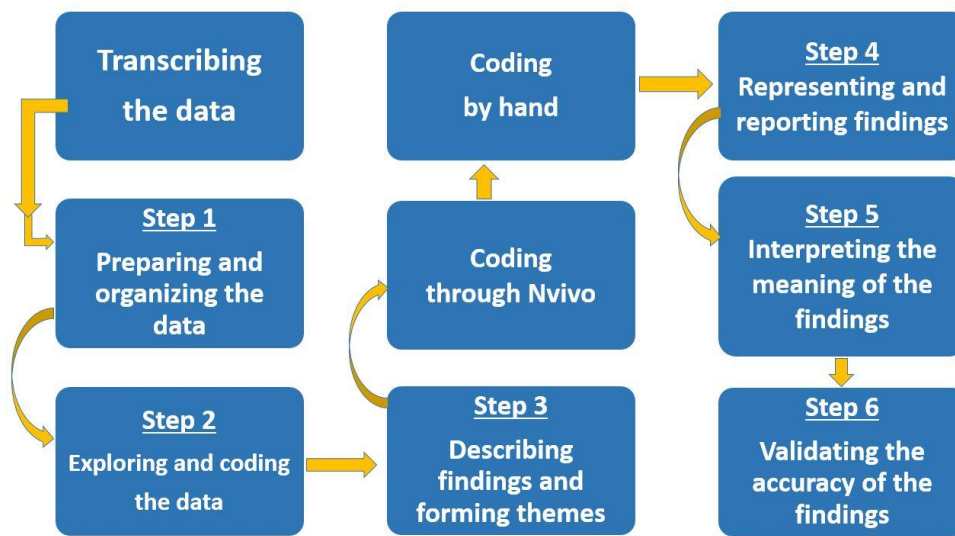


Figure 3.2. *The steps of the data analysis*

As Figure 3.2 presents, data analysis was conducted in six steps to ensure better quality of the research. Before the transcription process, the researcher adopted some strategies to ease the research study and increase the accuracy of the results. With this aim, a high-quality voice recorder was bought. In order to obtain a better-quality voice, the recorder was placed close to both the interviewer and the interviewees so that the clarity of the speeches would be achieved. The audio recordings of almost all the interviews, taken with the participants' permission, were transcribed through non-verbatim transcription. All the words stated by the participants were transcribed only excluding the pauses, fillers, stammers, and non-verbal dimensions. The process of transcription is claimed to differ depending on its use by the researcher (Mondada, 2007). The word "pause" was typed to indicate when interviewee(s) needed to take a break due to unexpected events such as phone calls, visitors, and urgent short leaves.

The researcher converted the words to a computer document, Word Document, as text data for the analysis of the data. When the recordings were vague, the researcher listened to the recordings several times and read his field notes to guess and complete the unheard part. The questions asked by the interviewer were highlighted in order to analyse only the answers. No transcriptionist was employed in this process as the researcher transcribed the information himself considering the confidentiality issues. At the end of the transcription process, 1.365-minute interviews with directors resulted in 663 pages (263.055 words) while 2.775-minute interviews with testing unit members ended up with 975 pages (382.707 words) of single-spaced transcription. The main points drawn from the transcripts were sent for a check to each university in order to ensure the validation of the interview results.

In the organisation and classification of the data, the researcher thought clearly about the format suitable for the printed transcription prior to the analysis process. Thus, a system of organisation was determined as the transcribed data was prepared and organized for analysis. The researcher considered the margin size and the amount of space to leave for comments, double spacing as 2-inch margins were created on each side of the text document for coding purposes, and extra space on the page between the interviewer's comments and the interviewee's comments so that he could jot down notes in the margins during data analysis.

The researcher conducted the initial analysis process by hand in order to make more sense of the text by immersing himself in the details and getting close to the data through a hands-on feel. Thus, to get a sense of the interview as a whole before breaking it into parts (Creswell, 2013), the researcher basically read the data and wrote some brief notes by highlighting key issues by hand as a first impression. He divided it into parts using colour coding to mark parts of the text, and cutting, pasting text sentences onto cards; then considered a large amount of database gathered during a study. The transcripts were read several times and examined critically considering the fact that each time the researcher read the data, he developed a deeper understanding of the responses supplied by the participants and a better relationship between what was mentioned during the interviews and the outcomes of this research.

While creating the patterns, similarity, and difference (codes emerging the same way or in a predictably different way), frequency and sequence (codes emerging often, seldom, or in a certain order), correspondence (codes emerging in relation to other

activities or events), causation (a code appearing to cause another) was taken into account. In this way, he realised that certain codes emerged in their responses (Saldana, 2011). Then the data was divided into codes as the smallest unit of data by evaluating and reviewing the emergent codes several times with the existing ones. Employing Excel, he developed a table of sources, used to help organize the material. By refining the codes through the analysis of each case, the researcher grouped the responses of the participants into categories and themes, respectively. In the end, a final table was developed to summarize the categories and themes along with the specific codes. Figure 3.3 presents a visual model adapted from Creswell (2013).

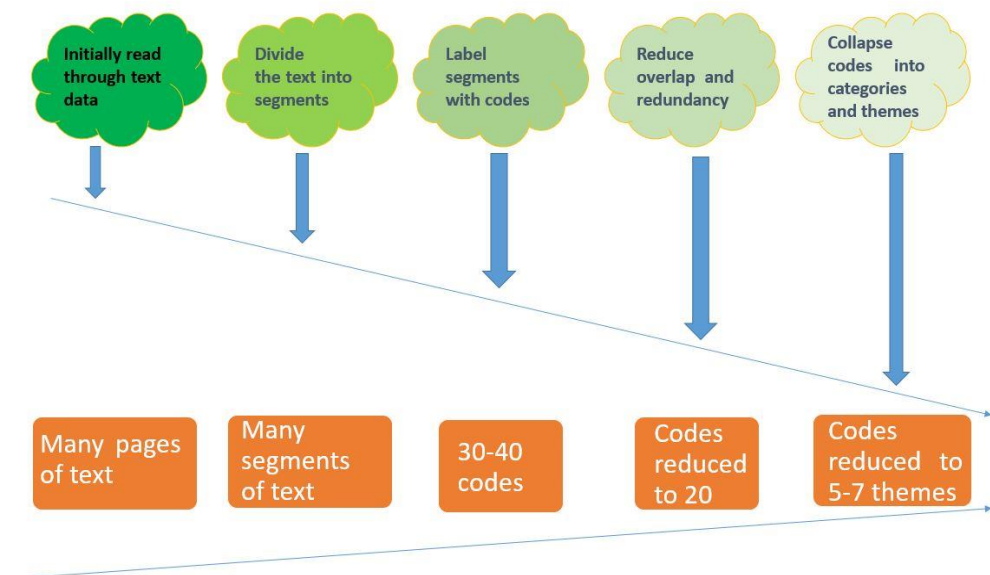


Figure 3.3. A visual model for the coding process

After initial coding, to easily keep track of files and locate text passages, the researcher used a qualitative computer programme, which is able to facilitate the process of storing, organizing, assigning labels or codes, analysing, sorting, and representing or visualizing the data. In this way, the researcher could search through his data and locate specific text or words. Nvivo was selected as it is easy to use and allows the researcher to read and review text and categorize it easily for the qualitative report. In order to employ the programme, the researcher got enrolled and has been following a training programme online (<https://www.iienstitu.com/>).

While refining the data, some codes were not merged on purpose so that they would not lose any stress and significance.

3.1.5. Validity and reliability

In the study, the accuracy or credibility of the findings was of utmost importance. The researcher should be self-reflective about his or her role in the research, how he or she is interpreting the findings, and his or her personal and political history that shapes his or her interpretation (Creswell, 2013). Thus, certain strategies were used to validate qualitative accounts in order to make sure that the findings and interpretations are accurate. Validating findings means that the researcher determines the accuracy or credibility of the findings through strategies such as member checking or triangulation (Klassen et al., 2012).

The interviews were conducted with both directors of language schools and testing unit members within the school, which ensured the triangulation of the data. Triangulation is the process of corroborating evidence from different individuals (e.g., a principal and a student) or methods of data collection (e.g., documents and interviews) in descriptions and themes in qualitative research (Miles et al., 2018). The inquirer examines each information source and finds evidence to support a theme. This ensures that the study will be accurate because the information draws on multiple sources of information, individuals, or processes. In this way, it encouraged the researcher to develop an accurate and credible report.

The researcher also checked his major findings with participants in the study to determine if his findings were accurate. Member checking is a process in which the researcher asks one or more participants in the study to check the accuracy of the account (Creswell, 2013). A summary of findings was sent back to the participants who were asked about the accuracy of the data in addition to many aspects of the study such as whether the description was complete and realistic, whether the codes and categories were accurate to include, and whether the interpretations were fair and representative. In this way, the participants were given a chance to change, clarify and elaborate more on what was said in the interviews. The findings were revised in accordance with the comments of the participants on the issues within the study.

Following the feedback of the interviewees, two independent raters analysed the data to assure the reliability of data analysis, in specific terms to check the accuracy of the accounts to ensure inter-rater reliability of the data. In addition to the researcher, an assistant professor working in an EPP of a state university, having written an MA

degree in testing, and having experiences with assessment practices and qualitative analyses was involved as a rater in the analysis of the interview data obtained from testing unit members. Besides, an assistant professor who was also vice director of an EPP in a non-profit foundation university having at least 15-year-experience of management in language schools was asked to take part as another rater in the analysis of the interview data gained from directors of language schools. Considering the substantial amount of data, the co-raters were asked to analyse 10% of the whole data collected through semi-structured interviews, which can be considered an adequate percentage to be used for inter-rater reliability considering the sample size and the number of codes and categories (Bernard, Wutich & Ryan, 2016; Lombard, Snyder-Duch & Bracken, 2004). Neither names nor any private data belonging to the participants were included in the transcriptions to keep their anonymity. First, 10% of the data were analysed and coded by the raters separately to form emerging codes, subcategories, and categories in relation to the purpose of the study. Upon the receipt of the coding of the colleagues, a negotiation session was conducted to review disagreements between the raters and finalize the codes. Creswell (2013) asserted member checking and inter-rater reliability make a great contribution to the accuracy and quality of the participants' comments and reflections regarding the results and findings of a study. To assess inter-rater reliability, Miles and Huberman (1994)'s percentage agreement formula was employed as shown below:

$$\text{Reliability} = [\text{agreement} / (\text{agreement} + \text{disagreement}) \times 100]$$

In this way, transcriptions were analysed by the researcher and two experts (one in terms of testing unit level and the other at the managerial level) to assure that the codes and categories are relevant making sense with regard to the purpose of this study. Inter-rater reliability was found as 92%. Since the inter-rater reliability was ensured, the rest of the data analysis procedure was conducted by the researcher.

3.2. The Second Phase

The methodology of this phase is based on the Delphi method, which is a qualitative method to gather and report group opinions. In this method, achieving a consensus on the common opinion is a key goal. Delphi, in most cases, is preferred

when there is a lack of previous research, and agreement, aimed to enhance the quality of decision-making (Avella, 2016; Powell, 2003). Besides, it is important to establish facts, generate ideas, or make decisions upon achieving a consensus on a broad issue (Blalock, 2014; Senyshyn, 2002). Delphi is described as “a democratic and structured approach that harnesses the collective wisdom of participants” (Powell, 2003, p. 381). A major factor is the anonymity of the panellists in preventing suppression.

There are mainly three types of Delphi in literature – classical, policy, and decision. However, combinations of several Delphis are used by the researchers as hybrid and modified ones. In this study, a modified Delphi approach was employed for gaining insight into common solutions in testing practices with regard to the challenges determined in the first phase of the study. The characteristic was ‘quasi-anonymity’ in which experts are called by name being known to each other from the beginning though the responses were anonymous (Hanafin, 2004; Shariff, 2015).

In education settings, Delphi is getting common in such studies as the development of standards and scales, education planning, and in other fields in which group opinion is of necessity in order to establish certain grounds (Korkmaz & Erden, 2014). The researcher considered using Delphi to receive solution offers in a quick and economic way (Nworie, 2011). The number of rounds and sample size differ in the literature. According to Skulmoski, Hartman and Krahn (2007), some studies might include one round, whereas others may consist of two or more rounds, which depends on the aim of the research.

In this part of the research, common opinions were attempted to collect and report to reach a consensus. Considering the fact that testing and evaluation in foreign language education is a long-lasting issue and its multidimensionality as well as its complexity, this method is selected as Delphi is chosen when there appears lack of previous research and knowledge, lack of agreement when the purpose is to raise the efficiency of decision-making (Powell, 2003) when the researcher aims to generate ideas, establish facts, or to make decisions and achieve consensus on a broad subject (Senshyn, 2002).

3.2.1. Participants

In this phase, seven experts were included in the Delphi method in order to obtain a consensus on the solutions. The participants in this phase were the ones having

experience in testing units at EPPs in higher education institutions. It is stressed that people have various perspectives of reality because of differences in their language, perception, and interpretation (Burton & Bartlett, 2020). In this way, experts in language testing could be involved to include various voices and reach common solutions. The participants work in different universities. The selection of participants for the Delphi method includes three criteria; 1-having worked as a testing coordinator, 2-having received intensive training on language testing, and 3-having at least 5 years of experience in preparatory schools.

The experts were informed briefly of the purpose of this study, their rights to withdraw at any stage, and a guarantee to keep their identity anonymous upon their demand. They were asked to sign an informed consent form, which covered the major information on what, how long, and for whom; their rights; possible benefits and the potential risks; confidentiality; and dissemination that should exist in an in-depth interview (Seidman, 2013). The language of the composition instruments was English as all of the participants were from the ELT department and could express their thoughts without language constraints.

The researcher did not experience any difficulty in selecting the appropriate panellists as he conducted the first phase to a broad extent, visiting 50 EPPs and ELT department if exists in each university. During the visits, the researcher was able to meet people who were interested in the subject under focus, share information about the study, and learn from them. The participants who volunteered to take part in this phase were contacted by telephone and e-mail and a virtual meeting was held via Zoom by appointment.

3.2.2. Data collection procedure

As mentioned earlier, in this phase, Delphi is employed so that language experts would attempt to find feasible solutions in accordance with the problematic issues determined in the first phase in terms of testing practices at EPPs in Türkiye. Since a consensus with sufficient information was reached, three rounds were conducted in this phase.

Modified Delphi adopted in this study includes applications similar to classical, policy, real-time, and e-Delphi. The rounds are explained below in detail.

3.2.2.1. Round 1: Written responses

The experts were sent the findings of the first phase to allow them to be aware of the challenges encountered in testing practices and asked to concentrate on finding solution offers accordingly. They were expected to write a set of draft offers without paying attention to their importance or matching them with the challenges, that is they started to list any ideas as solutions to gather their initial opinions. The essence of this round was to elicit initial opinions about common solutions in EPPs through free responses to a broad set of challenges.

3.2.2.2. Round 2: Focus group meeting

In this round, oral Delphi was applied online by means of two techniques: brainstorming and focus group interviews. The experts were invited to a zoom meeting, which, as there was a motivated group, avoided forcing them to travel from different cities, hosting the meetings, and allocating adequate financial resources. Before the meeting, the experts were provided with the initial list of ideas received via e-mail. The event started with an opening and welcoming part in which the experts were briefed about the previous steps of the study. Following this part, a brainstorming session was conducted as a warm-up in order to make them concentrate more and collect their ideas. This session as brainstorming was useful in breaking the ice in the expert panel meeting as there was a high participation rate.

The second part was organized as a focus group meeting in which the researcher elaborated on the ideas of the experts. Focus group meeting is described as a scientific data collection activity where ideas are reflected upon, stated, and perceptions or experiences of the participants are expressed in discussions in a non-threatening and free environment (Çokluk, Yılmaz & Oğuz, 2011; Creswell, 2013; Patton, 2014). In the focus group, the researcher collected the ideas and displayed them by means of online blackboard. The ideas were discussed and ranked in order to reach a group consensus. The lists of suggestions were revised, edited, and finalised for overlapping and repetitive suggestions, clarification of vague expression(s), the addition of new items thought to be essential, and deletion of unnecessary one(s). An assistant wrote in the list the solutions that were agreed upon by the experts. In other words, the consensus was achieved by discussion of the ideas of all panellists. No undesired negative effects such as status differences or peer pressure were experienced as the researcher provided the

freedom to each expert. The interactions in the panel are directed by the researcher himself. The event was a one-day activity, including two sessions, one in the morning and the other in the afternoon. Each session lasted for three hours.

The final lists of suggestions were reviewed by the experts taking part in the reliability of interview coding from the field of ELT. The ideas were checked in terms of language, relevance, and accurate grouping. Similar suggestions were merged, and the ones belonging to another category were transferred to the related list.

3.2.2.3. Round 3: Final approval of the experts

In the last round, the list of suggestions generated in the previous rounds was sent to the experts for their final approval. The ideas generated in the previous rounds were converted into statements in word format. Electronic confirmation was preferred since it was easier to deliver the list of ideas to participants without physical constraints. The experts were required to judge the common solutions determined in the expert panel and put a tick for the appropriate solutions.

In order to keep the response rate high, the researcher sent reminder(s) and made telephone call(s) when necessary. During this round, the participants may communicate with the researcher but not with each other in order to prevent the pressure from the experts with higher status through anonymity (O'Hara, De Souza & Ide, 2000; Powell, 2003).

3.2.3. Data analysis procedure

In the open-ended phases, the written data and interview data were analysed with content analysis and coding to determine themes, categories, and emerging codes. Content analysis is conducted with the coding of the data, read through as the meaningful parts are assigned codes with regard to the aim of the study (Powell, 2003). Based on the qualitative analysis of the responses given to each question, data reduction was achieved. The codes are merged into themes, categories, and subcategories, organized in a meaningful way in order to ensure that the relationships of the codes are identifiable (Yıldırım & Şimşek, 2006). Once the analysis of the data from interviews was completed, a cross-sectional analysis was performed that led to generalisations about the responses given to each question.

To conclude, this part of the study was an open-ended phase which consisted of the composition of written responses, an online panel which consisted of brainstorming and focus-group meeting, and a rating round. In the last round, the items were finalised by consulting experts.

3.2.4. Trustworthiness

The experts were sent the challenges found in the first phase to allow them to become aware of what is required to do. Then they were requested to submit their initial remarks and suggestions in the first step of the Delphi. In the next step, they were asked to join a Zoom meeting with a nickname to ensure the confidentiality of their names in order to avoid any hierarchy. This was intended to avoid any influence among them in terms of their status. In the last step, they were sent a draft of suggestions so that they could examine the results with a second eye. In this way, they had a chance to check the accuracy of the data in addition to many aspects of the study such as whether the description was complete and realistic, whether the codes and categories were accurate to include, and whether the interpretations were fair and representative, which ensured the triangulation of the data. The researcher also checked his major findings with participants in the study to determine if his findings were accurate. The findings were revised in accordance with the comments of the experts on the issues within the study. In this way, it encouraged the researcher to develop a more accurate and credible report.

3.3. Ethical Considerations

With regard to ethical considerations, permissions were obtained at different levels taking institutional review board approval, approval of language school, and individual approvals, namely from the directors and testing unit members before the researcher started to gather data. Beforehand, the researcher received Ethics Committee Approval from Anadolu University (see Appendix 1). To gain permission from organisational personnel, we contacted them before the start of a study and obtained their permission to enter and study their setting by asking for it formally in a letter noting the study was volunteer-based.

Considering the possibility that the university board and the directors of language schools may not be familiar with the specific topic, the letter included a detailed description and the procedures of the project, including the purpose of the study, the

amount of time he would be at the site collecting data, the time required of participants, and how he would use the data or results while stating the specific activities, the benefits to the organisation or individual because of the study, and the provisions he would make to protect the anonymity of study participants. The dates to conduct the interviews were determined by the participants taking into account their availability. One day before the interviews, the relevant participants were sent a reminder e-mail providing information on the general description of the study, purpose, estimated time, and interview questions so that they would reserve themselves for the interviews.

In this study, the names of the cities, higher education institutions, and participants are concealed. Some participants requested that the names of their universities not appear in the written documents. Thus, not only did the researcher have to conceal the names of the higher education institutions, but also the names of the cities in which the universities were located due to the fact that there was only one university in some cities.

Before the interviews, all the participants were informed that the anonymity, privacy, and confidentiality of the universities and specifically individuals who participated in the study were protected and the researcher was sensitive to the potential harm that participants might experience because of the study. It was ensured that participants consented to participation in the research study by developing an informed consent form for participants to sign before they participated in the study. A written consent was obtained from the participants even if the study posed minimal risk to the participants. This form stated that certain rights were guaranteed, including their right to withdraw at any time from the study, their voluntary participation in the project, and their right to know the purpose of the study and that when they signed the form, they agreed to get involved in the study and acknowledge the protection of their rights. They were assured that the data would not be shared with other participants or individuals outside of the study.

The procedures of data collection were administered with great sensitivity to the potential challenges and ethical issues of gathering responses face-to-face in their own environment. There was a balance between exploring the phenomenon, respecting the research environment, and disturbing or disrupting as little as possible. There were no potential power imbalances during the interviews, rather there was a friendly environment and the researcher was listening carefully to their endless challenges. They

were at a dead end. The anonymity of the participants was protected by masking the names of the individuals and the universities by assigning pseudonyms.

Bearing the research questions of the study, the researcher attempted to find answers to them by explaining the reason for the steps and making the link between the research questions and the individual phases and findings, as well. In the next section, the findings are presented.

CHAPTER 4

4. FINDINGS

Bearing the purpose of this study in mind, the challenges in the testing system at EPPs were examined through the analysis of data collected from interviews applied to both directors and testing unit members. Afterward, the solutions were searched through a focus group interview with volunteer experts in the field.

The chapter was divided into two major sections consisting of the results of the study in light of the research questions posed in the present study. First, the findings related to the first phase of the study, which were gathered through the interviews, are presented. The second section focuses on the findings from the second phase of the study as the solutions obtained through a focus group interview with field experts.

4.1. Findings of the First Phase

The aim of this phase is to examine the challenges encountered in testing practices at English language preparatory schools. To reach this aim, semi-structured interviews were conducted with directors/vice directors and testing unit members of the schools. The Constant Comparative Method was employed to analyse the data in a qualitative manner (see Section 3.5). The whole data was divided into emerging codes to find out the challenges in testing practices. The analysis revealed 17.253 codes in total.

Analysis of the qualitative data obtained through interviews with the participants (directors and testing unit members) indicated four main categories as the challenges experienced in testing processes. The main categories are given in Table 4.1.

Table 4.1. *Challenges in testing practices*

Categories	Number of Codes	%
Systemic	5.851	33,91
Stakeholder-related	4.804	27,85
Test processes	3.949	22,89
Other/general	2.649	15,35
Total	17.253	100

As Table 4.1. presents, the highest number of codes belong to systemic challenges consisting of national, institutional, and unit-related ones (N=5.851; 33,91%). The

second highest ranking category is stakeholder-related challenges consisting of directors/vice directors, testing unit members, instructors, and students (N=4.804; 27,85%). Test processes-related (N=3.949; 22,89%) and other/general (N=2.649; 15,35%) challenges are the last two categories. The Figure 4.1. illustrates the categories with their subcategories.

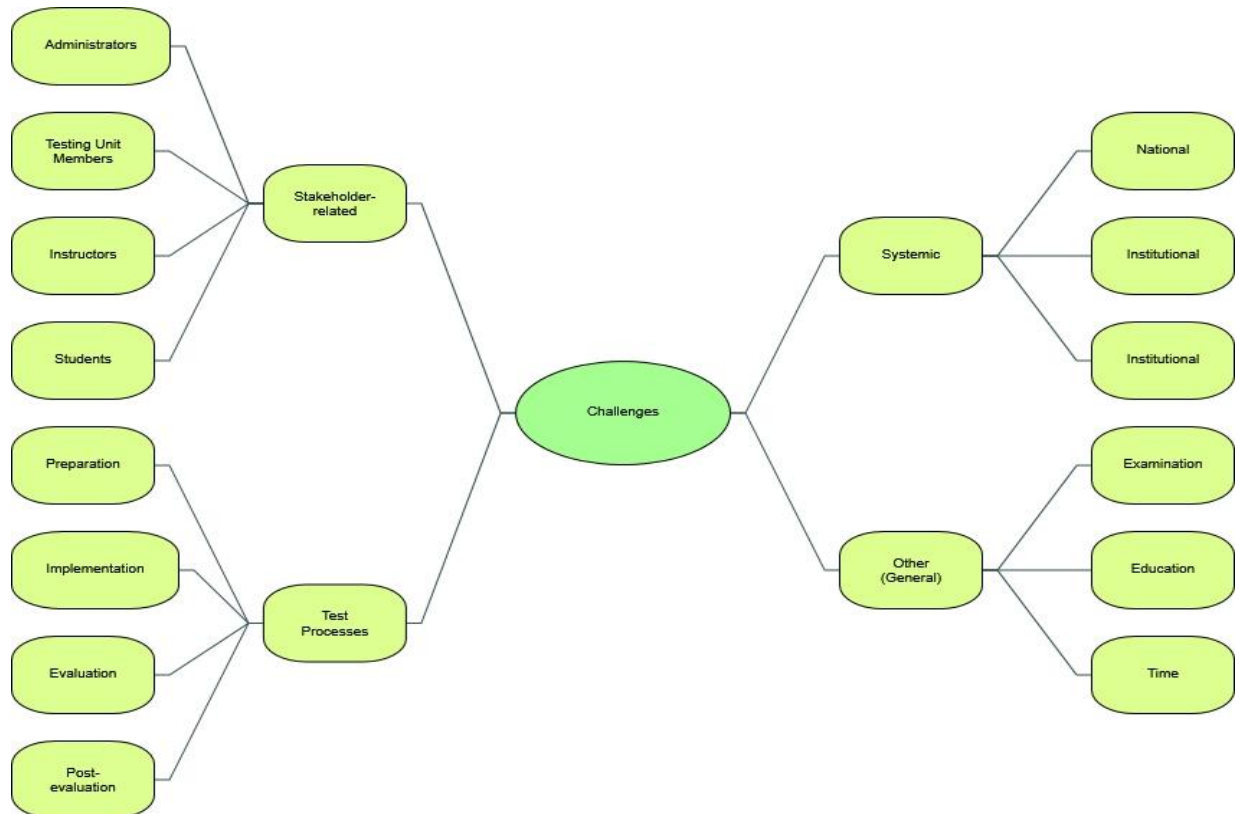


Figure 4.1. *The categories and subcategories of the challenges*

The remaining subsections explain each category and subcategory related to the challenges in testing practices in detail based on codes that fall into each category. The categories and subcategories are listed regardless of their frequency except for the codes. None of the categories existing in the literature were set as the basis for categorisation. Instead, emerging codes were compared and contrasted with each other by which the units showing similar characteristics were placed under certain categories and subcategories in line with the Constant Comparative Method. The categories and subcategories are presented with extracts from the interviews with the directors and testing unit members (see Appendix 5 for the list of codes in each category).

4.1.1. Systemic challenges

Systemic challenges comprise the most frequently stated type of challenges (N=5.851; 33,91%) encountered in testing practices at EPPs in higher education.

Table 4.2. *Systemic challenges*

Subcategories	Number of Codes	%
National challenges	2.033	34,75
Institutional challenges	1.998	34,15
Unit-related challenges	1.820	31,10
Total	5.851	100

As Table 4.2. presents, the highest-ranking subcategory belongs to national challenges (N=2.033; 34,75%). The second highest-ranking subcategory is institutional challenges (N=1.998; 34,15%). Unit-related (N=1.820; 31,10%) challenges are given in the last subcategory.

This category consists of national, institutional, and unit-related challenges. National challenges result from the decisions, actions, or legislative documents of authority government bodies, which seem to affect almost all testing practices. Institutional challenges are related to the applications, guides, and decisions taken by the senate, rector, or director within the university. The last one, unit-related challenges show the situations, organization, workflow of testing units at language schools.

In the following subsection, all the codes within this category will be explained in detail, respectively, by providing the comments of the directors and TUMs to support and show how the situation influence testing practices. The Figure 4.2. shows the subcategories and their codes of the systemic challenges.

As shown in Figure 4.2, national, institutional and unit related challenges are encountered as systemic challenges, which was complained frequently by the participants since the systemic challenges were claimed to limit the action plan of the university leaders or school directors and TU coordinators.

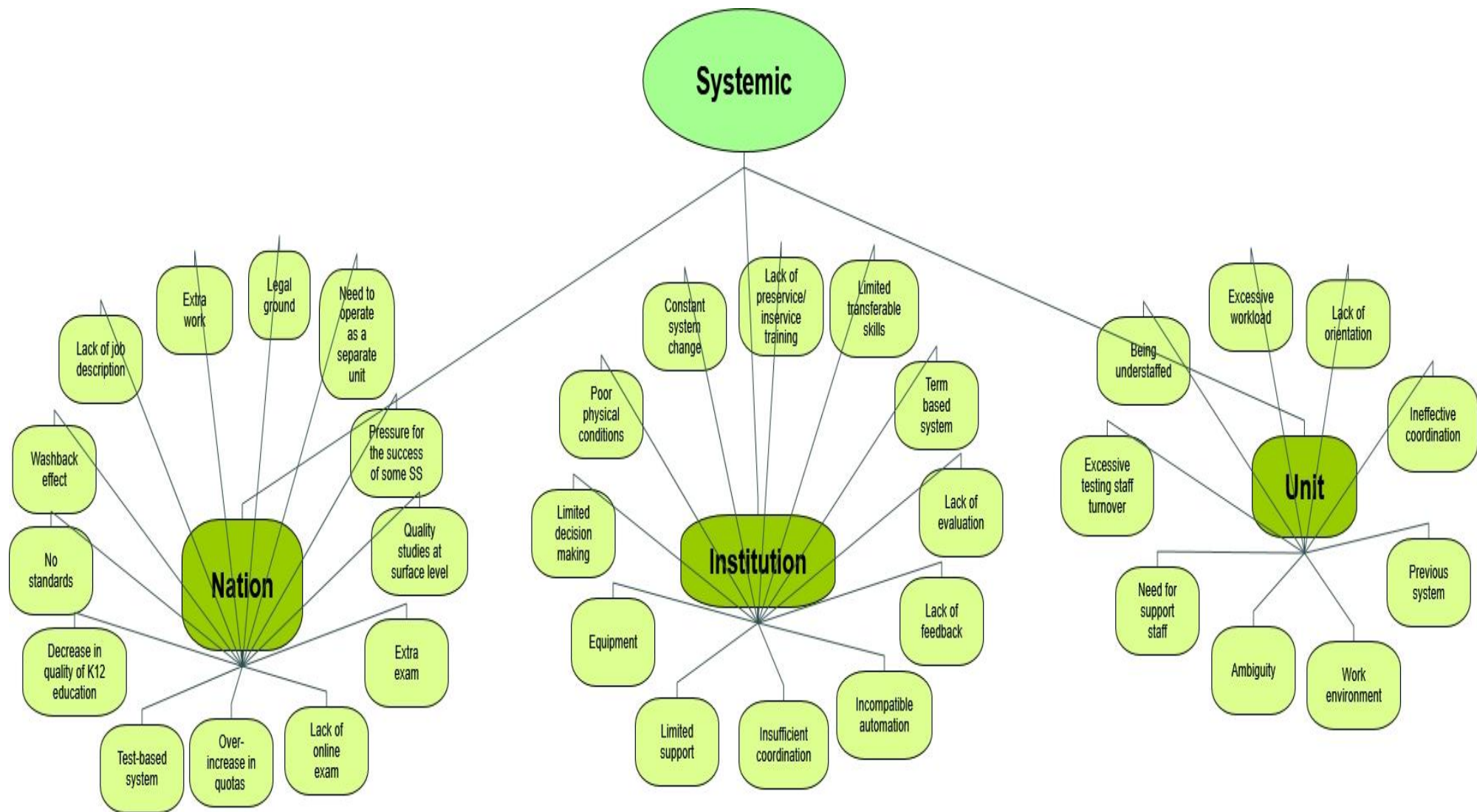


Figure 4.2. Systemic Challenges

4.1.1.1. National challenges

This subcategory includes thirteen emerging codes which are shown in Table 4.3.

Table 4.3. *Challenges at the national level*

Emerging Codes	Number of Codes	%
Legal ground	341	16,77
Uncompensated extra work	302	14,86
Lack of job description	269	13,23
Washback effect of YDS/YÖKDİL	231	11,36
No standards	169	8,31
Decrease in quality of K12 education	150	7,38
Test-based system	142	6,98
Over-increase in quotas	135	6,64
Lack of online exam	96	4,72
Need to operate as a separate unit	78	3,84
Pressure for the success of some students	52	2,56
Quality studies at surface or documentation level	41	2,02
Extra exam for late accepted students	27	1,33
Total	2.033	100

Almost all the interviewees indicated that the source of the challenges encountered in testing practices results from a legal loophole (N=341). Due to a lack of legal legislation, testing units are not recognized officially resulting in a lack of a title as TM, minimum workload, job description, and appropriate exit level. One testing member mentioned in the interview (All the extracts, including this one mentioned in the participants' native language, Turkish, were translated into English by the researcher; the italic ones in Turkish were written exactly in the participants' wording. The language of the interviews was Turkish but some participants used English terminology in the interviews, which is given in the original manner):

Aslında yasal zeminde böyle bir birim yok, bu açıdan baktığımızda testing member yani sınav hazırlayan diye böyle bir title da yok ama bizler testing ofis çalışıyoruz. Bunun sıkıntısını aslında çekiyoruz. Testing birimi ile ilgili yasal düzenlemeler yapılsa zorunlu sınırlar belirlenir. Bu kadar iş yükü belki gelmeyecek ve karşılığını da alacağız. Bu durumda biz bunun karşılığını alamıyoruz. Bu görevin yasal zemini olmadığı için yaptığımız işler ekstra bir külfet gibi algılanır oldu. Mesela nasıl müdürün görev tanımı var

görev tanımına göre iş veriliyor, belki görevin dışında illa ki iş yapıyordur ama en azından tanımı var. Tanımı olduğunda çok aşırı abartılmıyor. Sadece maddiyat da değil, orada müdür oluyor, müdürlükten başka bir yere yükselme şansı olabiliyor çünkü onun bir unvanı var. Ama bizim ofis işlerimiz bizi bir yere götürmüyor.

In fact, there is no such a unit on the legal ground, from this perspective there is no such a title as a testing member, but we are testing unit members. We actually suffer from this. If legal regulations are made regarding the testing unit, mandatory limits could be determined. Maybe much workload will not be given and we will get the reward. In this case, we are not paid back. Since this duty has no legal basis, the work we are doing here is perceived as an extra burden. For example, there is a job description of the manager, and the workload is given according to the job description, maybe he is working more, but at least he has a job description. It's not overly exaggerated. Not only financially, but he may also get a promotion after that position because he has a title. However, our office work doesn't take us anywhere.

One testing unit member in another university explained further:

Burada çalışacak kişiler eğitim alarak donanımlı, nitelikli hale gelse bile, hadi diyelim on kişi bile olsa burada şöyle de bir sıkıntı var, iki veya üç yıl sonra uygulamayı anlamaya başladıklarında, on kişinin gittiğini düşünün. Zorla verilen geçici görevlendirmeye stres altında, karşılığını almadan ne kadar dayanabilirler.

Even if the people working here become equipped and qualified by receiving training, and there are ten people, there is a problem here. When they begin to comprehend the implementation in two or three years, imagine that ten people are gone. How long can they stand under the stress of a forced secondment without being rewarded?

Another point was mentioned by a director:

Yasal açıdan Yükseköğretim Yeterlilikler Çerçevesinde ve hazırlık okulları yönergesinde çıkış seviyesi olarak belirlenen B2 seviyesi üniversitede kabul görmüş değil henüz. Diğer üniversitelerde de aslında merak ediyorum onlarda da öyle midir? B2 seviyesinin aslında olamayacağını düşünen çok büyük bir kesim var. Hatta sekiz ayda B1 bile çok imkanı varmış gibi görünmüyor, biz ne kitap kullanırsak kullanalım. Boş gelen öğrenciyi sekiz ayda nasıl çıkaralım o seviyeye. Bu konuda yasal çalışma yapılması lazım.

B2 level, which is legally determined as the exit level in the Higher Education Qualifications Framework and the preparatory school directive, has not yet been accepted within the university. I wonder if it is the same at other universities. There is a very large group of people who think that B2 level cannot actually be achieved. Even B1 level doesn't seem possible in eight months, no matter what book we use. How come we can raise to that level in eight months the students who come empty? Legal change should be done on this issue.

Another major concern of the directors and especially testing unit members related to national challenges is uncompensated extra work (N=302). Testing unit members are unhappy with being unpaid although they work under hard conditions and take some of their work home stealing from their personal time. Therefore, they are unwilling to continue to work in these offices. One of the testing unit members explained as follows:

Maddi motivasyonumuz yok her şeyden önce. Bazı kişiler sadece maddiyatı beklemiyor ama manevi kazanımların olmadığı durumda maddiyat da şöyle güzel oluyor; aldığımız cüzi bir ücret bile olsa en azından uğraşımızın bir karşılığı olmuş oluyor, "Ben niye yaptım" sorusu en azından bir cevap buluyor. Ek diye bir şey yok. Ofis çalışanları olarak kamu yararına çalışıyoruz gece yarısına kadar.

First of all, we have no financial motivation. Some people do not expect only financial support, but in the absence of moral outcomes, material goods are good; even if you get a limited fee, at least your effort is rewarded. The question "Why did I do it" at least finds an answer. There is no such a payment. As office workers, we work for the public good until midnight.

Another testing member explained further:

Burada herkes kendinden bir fedakarlık vererek özveriyle çalışmak zorunda, ama o fedakarlığın karşılığı olarak ekstra bir şey almıyoruz. Unvan yok, para istiyorum bari. O da yok. Para için demiyorum sadece. Hocalara göre biz daha fazla çalışıyoruz bence, doğal olarak daha fazla para almak zorunda gibi hissediyorum ben açıkçası. Çünkü ben derse giren hoca gibi 25 saat işimi bitirip gitmiyorum. Ben burada ekstra mesaiye kalıyorum ve bunun yetişmemesi benim suçum değil. Hiçbir şekilde kendi suçum olduğunu düşünmüyorum. Bir de hocalar 25 saat derse girdiğinde ek ders ücreti almış oluyor. Biz de 25 saatten fazla çalışıyoruz, evdeki çalışmalarınız hariç.

Everyone here has to work with self-sacrifice, but we don't get anything extra in return for that sacrifice. No title! At least I want money. There's nothing. I'm not just saying this for money. I think we work more than the instructors, I naturally feel like I have to get paid more because I don't work 25 hours and leave like the instructors. I'm staying overtime here and it's not my fault for not finishing my work. I don't think it's my fault in any way. In addition, when the instructors attend the class for 25 hours, they receive an additional course fee. We also work more than 25 hours, excluding our work at home.

As a number of participants indicated in the interviews, their job description was not identified in the legislation and language school directives. This lack of job

description (N=269) seems to affect their motivation and work in the testing unit. Owing to officially unspecified responsibilities, TUMs claim to be forced to do instructors' work and duties. A testing member cited her concern about the lack of a job description:

Görev tanımımızı bilmiyoruz çünkü yasal açıdan yaptığımız işlerin bir tanımı bile yok. Testing ofisindeki görev sorumluluklarımız, görev tanımlarımız belli değil. Keşke bunlar net olmuş olsa.

We don't know our job description because there is no legal description of what we do. Our job descriptions and responsibilities in the testing office are unclear. I wish these were clear.

Another testing member added:

Yaşadığımız sorunlardan biri de bizim burada idari bir görevimiz yok. Bir hoca gelip bize sınavla ilgili istediği gibi çemkirebiliyor. Çünkü bize bir iş ve sorumluluk veriliyor ama yetki verilmiyor. Anlatabiliyor muyum? Dolayısıyla ne oluyor, sürekli bu işin acı tarafını yaşıyoruz.

One of the problems we are experiencing is that we do not have an administrative duty here. A teacher can come and shout at us about an exam. That's why we are given a job and responsibility, but not given authority. Am I clear? So what's going on here, we are constantly experiencing the bitter side of this job.

One of the underlying reasons why we cannot teach and learn English dwells on the washback effect of language exams such as YDS, YÖKDİL (previously called KPDS or UDS). Some certain skills are not included in those exams, which affects learning and teaching habits. In the interviews, testing methods were mentioned to be influenced by the washback of YDS and YÖKDİL (N=231). One testing member expressed her feelings about the washback of such exams in these words:

Proficiency sınavımız YDS temel olarak alınmış bir sınava dönüştü. Maalesef amacımız sadece sınav yetişsin, kazasız yapalım şeklinde kestirip atıyoruz. Bu şekilde olmamalı diye düşünüyorum. Aksi halde tipik ÖSYM sınavı gibi oluyor ve mağduriyetler ortaya çıkıyor.

Our Proficiency exam has turned into an exam based on YDS. Unfortunately, our aim is just to get the exam done without any problem. I don't think it should be like this. Otherwise, it looks like a typical ÖSYM exam after which grievances arise.

A director added:

Hazırlık okulları dershaneye dönmekte, daha mekanik oluyor. Öğrenciler test sistemine alışık oldukları için genel olarak dili öğrenmek değil de sınavdan geçmeyi amaçlamaya başladı. ÖSYM'nin yaptığı sınavlara öğrenci yetiştiriyor gibiyiz, öğrenciler bizden onları teste hazırlamamızı istemeye başladılar. İşte skill geliştirmek, İngilizce öğrenmek yerine amaç “bana sınavı geçirt”e dönmekte. O kadar çok sınav odaklı ki öğrenciler.

Preparatory schools are becoming more mechanical than private language teaching institutions. Since the students are used to the test system, they have started to aim to pass the exam, not to learn the language in general. We seem to be educating students for the exams held by ÖSYM as students started asking us to prepare them for the test. Instead of developing skills and learning English, the goal has turned to “help me pass the exam.” Students are so exam-oriented.

There are 207 higher education institutions in Türkiye, which vary greatly from one another. Language schools in each university may also differ from each other. Although the differences may be beneficial from some points, standardisation becomes really tough, which causes different problems related to quality, outcomes, comparability and systematic issues, etc. The lack of standardisation (N=169) is one of the main challenges of testing members as indicated in the following extracts:

Avrupa Birliğinde dil becerileri açısından birlik sağlanmaya çalışılırken ülkemizdeki üniversitelerde çok farklı eğitim kalitesi var. Dil açısından aslında ölçütlere baktığınızda B1'in ne ifade ettiği minimum seviyede belli olması lazım. Ancak üniversitelerin eğitimleri, sınavları, sonuçları o kadar değişiyor ki. Birbirleriyle uyuşmuyor. Üniversiteler arasında çok eşitsizlik var. Mesela biz gramer odaklı sınavlar hazırlıyoruz ama başka bir üniversitede gramer sormuyorlar, onlar tamamen skills based sınavlar hazırlıyorlar. Soru tipinden diğer detaylara kadar çok farklı, zaten sizde incelemiştinizdir. Bazı üniversiteler listeninge daha az önem veriyor, diğer üniversiteler writinge daha çok önem veriyor. O yüzden nasıl olacak. Eğitimde ve proficiency sınavlarda ortak bir curriculumun olması gerekir.

While unity is being targeted in language skills in the European Union, universities in our country have very different education qualities. When you look at the language criteria, what B1 means should be clear at a minimum level. However, the education, exams, and results within universities vary so much. They do not match each other. There is a lot of inequality among universities. For example, we prepare grammar-oriented exams, but others do not include grammar, instead prepare completely skills-based exams. It differs in question type and other details, as you have already examined it. Some universities place less emphasis on listening, while other universities place more emphasis on writing. How is it possible? There should be a common curriculum in education and proficiency exams.

Mesela bizden önceki grupta bir süre sınav komisyonundakiler hiç derse girmemiş sadece sınav hazırlamışlar. Sonrasında beş saat ders yükü verilmiş, sonrasında sekiz saate ve bu sene 12 saate çıktı. Biz geldiğimizde ilk dönem mesela bizim ders programımız 12-16 saat olarak ayarlandı. Dolayısıyla bir sistemsizlik var bu konuda. Üniversiteler arası da ciddi farklılıklar var. Hem üniversite içerisinde kişiler değiştikçe hem de üniversiteler arası farklı uygulamalar bulunmakta. Aslında sınav birimlerinin görevlendirmesinde, uygulamalarında standart bir çerçeve belirlenmesi çok iyi olacaktır.

For example, those having worked in the testing commission before us did not lecture for a while, and just developed an exam. Afterward, a course load of five hours was given, then it increased to eight hours and this year to 12 hours. When we started, we were given 12-16 hours in the first semester. Therefore, there is a lack of a system in this regard. There are also significant differences among universities. There are different applications both within the university as administrators change and amongst universities. In fact, it is crucial to set a standard framework for the assignments and practices within examination units.

A testing member drew the attention to another side of the issue:

Bir öğrenci mezun olduğu üniversitede seviye olarak B1 çıkıyor ama atıyorum belki de x üniversitesinde A2 olarak çıkıyor, veya başka bir yerden B2 düzeyinde çıkıyor. Bunu nasıl o zaman standart hale getireceğiz. Bir de şunu söyleyeyim kimse kendisini kötü, kötü değil de aslında eksik göstermek istemez ama farklı üniversitelerde farklı seviyede sonuçlar çıkıyor. Bir kişi gelişmiş üniversiteyi mesela 200 puanla kazanıyor ama bizim üniversiteyi 180'le kazanabiliyor, ya da başka üniversiteyi 150 ile kazanıyor. Bu üniversite iyidir, diğeri kötüdür şeklinde karşılaştırmayı halk yapabilir sorun değil. Ama önemli olan üniversitenin de kendini geliştirmesi gerekir. Biz hep maalesef negatif açıdan bakıyoruz ama bazı noktaları benchmarking ile daha iyiye götürebiliriz. Kağıt üzerinde B1, A2'den ziyade akademik çalışmalarla uygulama düzeylerini ve sınavlarımızı benchmark edebilsek, başarıyı yakalarız. Sonuçta amacımız dil eğitimi.

A student exits as B1 from a university, but as A2 from x university, or B2 level from another one. How can we standardize this then? Let me also say this, no one wants to show themselves as bad, actually not bad, but deficiencies, but different universities have different results. For example, a person can get enrolled at a developed university with a score of 200, but at our university with 180, or at another university with 150. The public can make the comparison that this university is good while the other is bad. It is normal. But the important thing is that the university also needs to develop itself. Unfortunately, we always take into account negative points, but we can improve some points with benchmarking. If we can conduct benchmarking of our applications and exams through academic studies rather than labels such as B1, A2, we can catch success. After all, our aim is language education.

One director explained how integrating the students having completed EPP at another university causes anxiety in the following extract:

*Şimdi bize B*** Üniversitesi'nden yatay geçişle gelen öğrenci var, orada hazırlık okumuş ama ben oraninkini kabul edemiyorum. Kimse kimseninkini kabul edemiyor, çocuk diyor ki "hocam ben B***'de okudum," tatlım "tamam, ama maalesef bize YÖK tarafından tanınan bir sınavdan puan getirmen lazım ya da bizim muafiyetten geçmen lazım" diyorum. Öğrenci haklı ama ben de bir şey yapamıyorum. Geçerli puanı yoksa benim sınavımı almak zorunda. Keşke sınavlarımızın en azından ülkemizde geçerliliği olabilse.*

Now, there is a student who came to us from B*** University and had preparatory education there, but I cannot accept it. No one can accept a score taken from another university, the child says, "I studied at B***," and I reply "Ok, but unfortunately you have to bring us a score from an exam recognized by YÖK or you have to pass our exemption." The student is right, but I can't do anything. If he doesn't have a valid score, he has to take my exam. I wish our exams could be valid at least in our country.

A director warned:

Tepe kurumlar bir standart belirlemedikçe üniversitelerde kişiler, yöneticiler değiştikçe farklılıklar ve boşa uğraşlar devam edecektir. Biz görece geldiğimizde süreçler belli değildi, kimin sınav hazırladığı bile değildi. Bir yapı oluşturduk ancak bizim görev süremiz bittiğinde, bayrağı bizden teslim alan bir sonraki yönetici buna dikkat etmezse tüm bu gelişmeler çöp olabilir. Çünkü gereklilik ve zorunluluk olmadığında sistematik değil bu çalışmalar.

Unless authority institutions set a standard, differences and wasted efforts will continue as people and administrators change in universities. When we started, the procedures were not clear, it was not even clear who prepared the exam. We have set a structure, but all these developments may be garbage if the next manager who takes over the flag from us does not pay attention. That's why these studies are not systematic when there is no necessity or regulation.

In addition to the challenges stated above, the decrease in the quality of K12 education (N=150) is a key factor contributing to lowering the standards among universities. Although students start to receive education at a very young age, namely at the age of 7, language education during K12 still has not reached the desired level. Thus, post-high school students are believed to have low background knowledge in English when they enter higher education. Moreover, the TUMs interviewed surprisingly mentioned that most of the new students start preparatory school with an A level. One of the testing members specified her concern on this issue as follows:

MEB o kadar bozuldu ve gelen öğrenci profili o kadar düştü ki hocam. Geçenlerde haberlere bile çıktı, öğrenciler 2-3 matematikle mühendisliğe gidiyor, böyle bir rezalet olabilir mi? Mühendislik ya adı üstünde matematiğinin iyi olması lazım 2-3 matematikle artık mühendislik okuyanlar var. Bu öğrencilerin dile verdikleri önemi ve dil seviyelerini hiç yadırgamıyorum.

MEB has deteriorated so much and the incoming student profile has dropped so much. It's even been on the news recently students get placed in engineering with a 2-3 maths score, how disgraceful it is! The students studying engineering, on the other hand, must be good at mathematics. There are now those who study engineering with 2-3 maths score. I don't find strange the importance that these students attach and their language level.

A director drew the attention to another side of the issue:

Hocam bazı şeyleri değiştiremiyoruz çünkü önce kendi başarımızı sorguluyoruz sonra dönüp dolaşıp K12'deki yetersiz ve daha bozulan eğitim kalitesine çıkıyoruz. Bize gelen öğrenci profili o kadar düşük ki. İdeal olarak keşke daha iyi gelse öğrenciler, ben hep diyorum yani ortaöğretimden öğrenci hazırbulunuşluk düzeyi yüksek gelse biz de onlara burada akademik İngilizce üzerine gelişmelerine yardımcı olabilesek. Bölüm İngilizcesi vereyim. Ama size gelen öğrenci sıfır düzeyinde yani eksi düzeyinde geliyor neredeyse çoğu üniversitede.

We cannot change something because initially questioning our own success, we conclude that the quality of education in K12 is deteriorating and getting inadequate. The incoming student comes to the university with a low profile. Ideally, I wish students would be better, I always say, if the level of students is high after secondary education, we can help them improve their academic English here. If only we provided them with technical department English. But the students start higher education at zero level, that is at the minus level, at most universities.

Another director clarified how this situation made him anxious:

Şu zamanda hala İngilizce öğrenmenin önemini anlatmaya çalışıyoruz ya, çok gülüyorum. MEB'de bu aşılsa, ilkokul 2. Sınıftan beri İngilizce dersi görüp hala İngilizcenin önemini anlamayan çok kişi var. Keşke sadece bu verilmiş olsa bırakın ileri İngilizce düzeyini.

At this level, we are still trying to explain the importance of learning English, I laugh a lot. If only this were given within MEB, there are many people who have taken English lessons since the 2nd grade of primary school but still do not know even the importance of English. If only at least this were given, let alone the advanced level of English.

Another code is the test-based system (N=142). Although it is emphasized that the participants were well aware of the importance of language skills, they transformed

their education into test-oriented excluding skill-based exams due to the reasons they could not change. When preparatory education is test-oriented, the focus on language learning might be lost. Instead of learning in the process, students pay attention to grammatical information with mechanical question structures and formulations in order to get prepared for the exam and eventually pass it. Besides, the number of exams pushes TUMs who work in the office as understaffed, which is believed to cause time constraints, stress, and demotivation. One testing member expressed her feelings about the test-based system in these words:

Sıklıkla sınav yapıyoruz çünkü bu sınavların verilmesi gerekiyor diye yönetimden talep geliyor ve bu sınavlar veriliyor. Sonuç olarak öğrenciler geçti kaldı, onun dışında başka bir değerlendirme yok dil edinimi gibi bir amaca yönelik bir şeyimiz yok. Öğrenciler mesela quiz var diye çalışıyorlar. Bu hiçbir zaman engelleyebildiğimiz bir durum değil. O yüzden de sıklıkla sınav yapıyoruz.

We often make exams because there is a demand from the management that these exams need to be given. As a result, the students either pass or fail, we do not have any other evaluation, even if we don't aim at language acquisition. For example, our students are studying because there is a quiz. This is not something we can ever prevent. That's why we often test.

One testing member was anxious about test-based education and she explained the situation in the interview as follows:

Hem üniversitede aldığımız eğitimler hem de tecrübelerimiz göz önünde bulundurulduğunda dil öğretiminde beceri odaklı eğitimin ve nihayetinde sınavların önemi aşikar. Ancak sayımızın az olması, sınavlarımızın high stake olması ve iş yoğunluğumuzdan dolayı beceri odaklı sınavlardan ziyade daha mekanik soruları ve sınavları tercih etmeye başladık. Kağıtları okuyan hocalar da memnunar. Öğrenciler de genetic olarak sınav bazıli sisteme yatkın oldukları için beceri geliştirmeyi dert etmiyorlar.

The importance of skill-oriented education and the following exams is obvious in language teaching considering the education we received at the university and our experience. However, due to the fact that we are understaffed, that our exams are high-stake exams, and that our workload is much, we started to prefer more mechanical questions and exams rather than skill-based exams. The teachers who score the papers are also satisfied. Since students are genetically prone to the exam-based systems, they do not worry about developing any skills.

Furthermore, the over-increase in quotas (N=135) is a noticeable concern of the testing members. The inverse rate between staff shortage and student quota seems to affect the examination process. During the interviews, it was frequently stated that decisions and policies of governing bodies having a lack of coordination contradicted real situations and the state of EPPs. Therefore, an over-increase in student quota is felt to disturb a number of students and classroom balance even though a new instructor is rarely appointed to preparatory schools. One testing member stressed the reason for the crowd of students:

Öğrenci çokluğunun ana kaynağı her sene kontenjanın artması.

The main source of the student population is the increase in the quota every year.

A director explained the situation as follows:

Ben ulusal otorite niteliğindeki YÖK'te, YÖKAK'ta açıkçası böyle bir kopukluk olmasına bir anlam veremiyorum. YÖKAK'ın yaptığı değerlendirme toplantısında biz uzun uzadıya sınıflarımızdaki öğrenci sayısının 42 olduğunu ve bu sayının dil eğitimi olumsuz etkilediğini anlattık. Bunun çok açık bir sebebi var; YÖK'ten kadro alamadığımız gibi öğrenci kontenjanı her geçen yıl katlanıyor. Daha komiği ise bize gelen raporda sınıfları ve ders materyallerini daha verimli kullanabilmek için sınıf sızelerini düşürün. İnanabiliyor musunuz bize classroom sızeleri düşürün dendi. Ülkemizde denetleme ile ilgili bir şeyler yapılıyor. YÖK bir şey yapıyor, YÖKAK bir şey yapıyor, ÖSYM bir şey yapıyor, MEB bir şey yapıyor. Ortaya bu çıkıyor. Aralarında koordinasyon yok.

Frankly, I can't understand why there is such a discoordination between the national authorities such as YÖK and YÖKAK. At the evaluation meeting held by YÖKAK, we explained in detail that the number of students in our classes is 42 and that this number negatively affects language education. There is a very clear reason for this. On one hand, we could not get an instructor from YÖK; on the other hand, the student quota is increasing every year. What's even funnier is that in the report sent by YÖKAK, we were asked to reduce the class size in order to use the classes and course materials more efficiently. Can you believe this? We were told to reduce classroom sizes. In our country, something is being done about evaluation. YÖK is doing something, YÖKAK is doing something, ÖSYM is doing something, MEB is doing something. What turns out is that there is no coordination between them.

Another challenge experienced by the participants was reported as a lack of online exams (N=96). COVID-19 brought new challenges in language education and testing. Although technology has advanced a lot, online exams were not favoured in

proficiency exams by preparatory school administrations earlier than the pandemic. It was interesting that when digitalisation and online testing were asked during the interviews, TUMs told that digitalisation was included merely within tasks. One testing member cited her concern about this lack:

Online sınav yapılmıyor. Online materyaller kullanılıyor ama bizim burada yaptığımız online bir sınav yok. Yalnızca, bir kur içerisinde öğrencinin geçmesine etki eden ölçme araçlarından %5'lik olanı online bir task.

There is no online exam. Online materials are used, but we do not have an online exam here. Only 5% of the evaluation tools is an online task.

A testing member focused on the lack of online exams although they provided online education and online exam under distance education:

Online sınavımız yok. Aslında uzaktan eğitim kapsamında İngilizce sınavları yapılıyor, ama hazırlık bölümünde direkt online sınav yapmıyoruz.

We do not have an online exam. Actually, there are English exams within the scope of distance education, but we do not conduct direct online exams in preparatory schools.

Another emerging code is the need to operate as a separate unit (N=78). The name of the testing unit seems to change from one university to another, yet the unit has neither power nor significant importance no matter which name it has. Therefore, it is needed to transform the unit into a legally separate unit which has its own financial budget under the school. One testing member who was anxious about this need mentioned in the interview:

Daire ya da birim gibi tanınmıyor aslında mevzuatta. Daire ya da birim olmayınca siz aslında görevlendirilemiyorsunuz yasal olarak. Bu da aslında uygulamada ihtiyaç duyulan bir şey. E o zaman, yasal zemine niye kavuşturulmuyor.

In fact, it is not recognized as a department or unit in the legislation. Since it is not a department or unit, you cannot be assigned legally. This is actually needed in practice. Then, why doesn't it have a legal basis?

A testing member added:

Ayrı bir oda veya başkanlık olsak, ayrı ödenek olur herhalde. Fiziksel koşul olarak baktığımızda bile biz bir oda veya başkanlık olsak neden mesela fotokopiyi kendimiz çekelim veya onları kendimiz taşıyalım. Bütün diğer başkanlık ve odaların zaten bunların memurları da var, ulakları var.

If we were a separate unit or department, we would probably have a separate budget. Even considering physical conditions, why should we make photocopies on our own or carry them if we were a separate unit or department? All other units or departments already have officers and carriers.

Another point made by the participants was their concern about the pressure for the success of some students (N=52). One testing member who was concerned about the pressure thought that no matter how objective and perfect exam experts prepared, some external interventions could not be avoided. In addition to students themselves, their family members or relatives may tend to get in touch with the rectorate or directly with preparatory school management to ask for the success of their student, which might also come as an order from bureaucrats in some situations. The following extract explains his feelings:

İdareden bazı öğrenciler için “şu geçsin, bu kalsın” şeklinde telefonlar geldiğini duyuyoruz. Yönetiminden de desteğiyle biz dikkate almıyoruz.

We hear that our administrators receive some calls regarding some students as “this student will pass, that student will not.” With the support of our administrators, we do not take the calls into account.

A testing member explained her stressful moments related to the external pressure in these words:

Bazı öğrencilerin anneleri, babaları da ofise geliyor zaten. Ben savcıyım, ben avukatım benim amcamın oğlu hakim falan. Milletvekillerinin idareyi aradığına şahit oldum.

Some students’ mothers and fathers come to our office. We are told “I am a prosecutor, I am a lawyer, my uncle’s son is a judge” or something. I witnessed MPs calling our administrators.

Quality studies at the surface or documentation level (N=41) also affected the motivation of testing members in a negative way. Quality assurance studies have gained momentum in Turkish higher education with the establishment of the Higher Education Quality Council. However, especially in terms of EPPs, quality-oriented evaluation and studies remain at the surface level. Thus, all the work requested from TUMs is believed to be extra workload and time-wasting when preparing documentation in a hurry schedule only during site visits of governing body experts. One director clarified how these studies affected testing members:

Kalite kültürünü oturtmadan, kalite kavramına alışmadan, kurum ve çalışanları bu nosyonu benimsemeden kalite oluşturmak zordur, hatta imkansızdır açıkçası. Ya da kalite kültürünü sağlamış bir üniversitenin sistemini başka yere uygulamak da mantıklı değil. Eğer kalite kültürü içselleştirilmezse yapılan işler ve süreçler formalite boyutunda kalır. Kalite belgesi levhasını bir yere asmakla kalite sağlanmış olmaz.

It is difficult, even impossible, to create quality unless the institution and its employees establish a quality culture and adopt the concept of quality, the notion. It is also not logical to apply to another place the system of a university that has established a culture of quality. If the quality culture is not internalized, the work and processes remain as a formality. Quality is not ensured by hanging the quality certificate plate somewhere.

A testing member exemplified how they were stressed out:

Kalite çalışmaları başladığından beri aşırı bir belgeleme oluyor. Bütün sınav arşivini istiyorlar, specificationları istiyorlar, sürekli her yaptığımız şeyi kağıt olarak raporlamamız isteniyor. Mesela şu an bizden en yoğun zamanımızda bu süreçleri fotoğraflamamızı bile istediler. O dereceye kadar, düşünün her şeyin belgesini istiyorlar, aklınıza gelebilecek her şey. O da ekstra bir yük, ayrı bir sekretarya işi bizim için ve peyderpey gelip söylüyorlar. Toplantı programları belirsiz, önceden bilsen sen ona göre planlarsın. Okula her geldiklerinde toplantıda “analiz yaptık, şu eksik” şunu yapın!” diyorlar. Onlarında toplu bir listesi yok sanırım. Vardır belki de vermek istemiyorlardır ya da bunu bir sürece yaymak istiyorlardır. Müdür bey bizden bir belge istedi ve biz aşırı doluyuz şu anda, şimdi sınavı bırakıp onları yapmak zorundayım ben bir yandan. Şimdi ben bunu mu yapayım, onu mu yapayım oldu. Yine bir sekretarya işi baktığınızda ve sadece belge yükü. Dolayısıyla kalite süreci gelişmemize tam sirayet etmemiştir. Gerekçesi ile ilgili bize bir şey söylenmedi.

There has been excessive documentation since quality studies began. They want the whole exam archive, including the specifications, and we are constantly asked to report everything we do. For example, they even asked us to take the picture of these processes during our busiest time. They want documentation about everything. They say it's an extra burden, a separate secretarial job for us and they come gradually. Meeting schedules are uncertain, if we know beforehand, we can plan accordingly. Whenever they come to school, at the meeting they tell us to do the missing by saying “we did the analysis, this is missing.” I don't think they have a complete list. Maybe they have but don't want to share it with us or they want to spread it in a process. Our administrator asked us for a document but we are overloaded at the moment, now I have to leave preparing the exam and do what is asked. Now, should I do this or that? Again, it is a secretarial job and only documentation. Therefore, the quality process has not fully contributed to our development. We were not told anything about the reason.

A testing member drew the attention to another side of the issue:

Ülkemizdeki eksikliklerden bir tanesi de kalite ile bir şey değişecekse işin mutfağında çalışan kişilerden talep edilerek belirlenen ihtiyacın bir nevi düzeltilmesi gerekir. Buradaki iyileştirme çalışmaları ise yukarının belirlediği kıstasların sorgulanması şeklinde olduğunda benim gerçek ihtiyaçlarım giderilmiş olmuyor.

One of the problems in our country is that if something will change with quality, the need should be determined by the demand of the staff working in the kitchen so that the change could be corrected in an appropriate way. When the improvement work is conducted by only questioning the criteria set by the authority institutions, my real needs are not met.

The final code in this subcategory is that testing offices are to prepare an extra exam for late acceptance of international students or additional placements of Turkish students (N=27). The participants stressed that since there is no general systematic practice in most universities, foreign students are registered after general registration; thus, they have to prepare additional exams for these students. As they stated, a similar situation is experienced in the registration of the students who come with additional placement and the exemption exam that is conducted at the beginning of the year cannot be done in one go. One director clarified how these studies affected testing members:

Keşke tek bir sınav yapabilsek, daha iyi yapabilmek için uğraşırız. Ancak tam proficiency sınavını bitirdik diyoruz, yabancı uyruklular geliyor, o bitiyor, ek yerleştirmeler... Her bir grup için ayrı bir sınav yapmak zorunda kalıyoruz. Bu kayıtlar birbirine yaklaştırılrsa, tek bir sınav yapabilsek bizi bu kadar yormaz.

I wish we could only do one exam, we could try to do it better. However, we think that we have completed the process of the proficiency exam, then foreign students are registered. We complete it but then additional placements... We have to do a separate exam for each group. If these registrations were brought closer to each other, if only we could only do one exam. It wouldn't tire us that much.

National challenges seemed to have an impact on other challenges. The compensation of a national challenge was claimed to be rather difficult by the directors since they were not empowered with responsibility due to the lack of legislation. Besides, washback was seen as a major challenge influencing not only language testing but also language education at EPPs. Another key challenge was limited digitalisation. The next subsection deals with institutional challenges.

4.1.1.2. Institutional challenges

This subcategory includes twelve codes which are shown in Table 4.4.

Table 4.4. Challenges at the institutional level

Emerging Codes	Number of Codes	%
Lack of pre-service/in-service training	332	16,62
Constant system change	273	13,66
Poor physical conditions	254	12,71
Limited decision making	226	11,31
Problems with equipment	205	10,26
Limited financial contribution/support	179	8,96
Insufficient coordination among the units	157	7,86
Incompatible automation system	132	6,61
Lack of feedback system	107	5,36
Lack of evaluation system	64	3,20
Term based system	43	2,15
Limited assessment of transferable skills	26	1,30
Total	1.998	100

The most frequent challenge encountered in testing practices is mentioned as the lack of pre-service/in-service training (N=332). Without pre-service training, testing members feel lost when they are assigned to work in the office. What is more, they are rarely supported with in-service training especially in state universities due to the limited financial budgets of the schools. During the interviews, the directors who were even aware of the importance of training complained that they might not come up with a sustainable solution other than the ones provided by publishing houses in return for the use of their coursebooks. One director mentioned in the interview:

Bu ofislerde çalışanların eğitimlerinin olmaması büyük bir problem. Biliyorsunuz bu alanda çok fazla eğitimli insan yok. İşte bir elin parmağını geçmiyor nerdeyse. Eğitimli insan sayısının az olması en büyük problem. Üniversitelerin kadrolarında konu ile ilgili eğitimli kişilerin olmaması, olan bir kişi varsa da zaten onu hemen özel üniversiteler kapıyor ve direkt testing birimine görevlendiriyor. Hizmet öncesi eğitimi olan uzman veya üniversitelerin görevlendirme öncesi bir eğitim vermesi çok nadir. Hizmet içi daha çok, o da imkanlarla sınırlı. Aslında testing konusunda yetkin kişiler olması lazım, gerekli eğitimi almış olmaları lazım. İdare olarak daha fazla konuyla ilgili eğitim almaları için oldukça çaba sarf ediyoruz ama kaynak yetersizliğiyle ilgili bir durum da var. Eğitimi verebilecek gönüllü bir insan bulmak şu an kurumda güç, diğer yandan ise ücret ödemiyoruz. Artık

yayınevleriyle anlaşıyoruz. Eğitimler karşılığında tüm kitaplarımız anlaştığımız yayınevlerinden alıyoruz.

The lack of training for those working in these offices is a big problem. As you may know, there are not many educated people in this field. There are very limited people. The biggest problem is the low number of educated people. There is a lack of trained people at universities, and if there is a trained person, private universities immediately grab her/him and directly assign them to the testing unit. The number of experts who have received pre-service training or of the universities which gives training before the assignment is very limited. More as in-service training, which is also limited based on possibilities. In fact, testing members must be competent in testing, they must have received the necessary training. As administrators, we make a lot of effort for them to receive more training on the subject, but there is also a lack of resources. It is difficult to find a volunteer person who can provide training in the institution, on the other hand, we do not pay any fees. Then we are dealing with publishers. In return for training, we buy all our books from the publishing houses we have contracted with.

One of the testing members explained as follows:

Özellikle devlet üniversitelerinde item-writing kısmında daha çok yeni gelen arkadaşlar oluyor. Sistemi tam henüz kavrayamamış olduklarından, functional, practical değil de daha çok standart, gördüğünü ölçen sorular hazırlayabiliyorlar. Mesela eski spesifcationlara ve hazırladığım eski sınavlara baktığımda, çok hatamın olduğunu fark ediyorum, çünkü testing uzmanı olmadığım için bazı şeyleri ben de burada öğrendim. Mesela belli bir kelime sayısını, işte 150 kelime aştığında ne gibi problemle karşılaşabilirim bunu öngöremiyordum veya neden tam olarak o kelime aralığı bilmiyordum. Altındaki teoriği tam bilemeyince, ne kadar önemli diyorum kendi kendime. Mesela bir reading de 1000-1200 kelime arası diyor benimki 1350. Çok büyük bir sorun mudur bilmiyordum. Analizlere bakmak konusunda eksik kalabiliyoruz. Örneğin ben yeni bir metin yazdım spesifcationda, dip barajı 80 olacak diyorum, benimkinin 60 çıktı, bunu 80'e nasıl çıkarabilirim gibi sorunlar.

Especially in state universities, the new instructors are assigned to conduct item writing. Since they have not fully grasped the system yet, they can prepare more standard questions that measure what they see, rather than functional or practical. For example, when I looked at my previous specifications and exams, I realized that I made a lot of mistakes because I was not a testing expert, so I learned some things here. For instance, I could not foresee what kind of problem I might encounter when a certain number of words exceeded 150 words, or I did not know why exactly that word range. When the theory is not known, I wonder myself how important it is. For example, a reading requires between 1000-1200 words, mine is 1350. I didn't know if it was a big problem. We lack follow-up analyses. For

example, I wrote a new text in the specification, it says the bottom threshold will be 80, mine is 60, and how can I increase it to 80?

Another testing member added:

Eğitimle ilgili sıkıntılar, genellikle alamadığımız eğitim daha doğrusu. Hizmet öncesi eğitim almadan ofiste çalışabiliyoruz. Ofis içindeki daha önceden çalışıp, bilen insanlar kendi yoğunluklarından her zaman yeni gelen kişiye yardımcı olamayabiliyor. Çok strict olabiliyorlar, çünkü başkasına yardım etmek ve anlatmak da zamandan çok çalan bir şey. Dolayısıyla, bizim öğrenmemiz için bir oryantasyon veya hizmet öncesi program yok. Yani yaparken aynı anda öğrenmemiz gerekiyor. Yardım istediğimizde bilen arkadaşlarımızın zamanından çok fazla çalışıyoruz. Zaman yetersizlikleri sebebiyle de ek işler için kişisel zamandan harcamak durumunda kalıyoruz.

Problems with training, or rather the training we usually do not receive. We can work in the office without pre-service training. People in the office who have worked before and know the system may not always be able to help the newcomer due to their own intensity. They can be very strict because helping and explaining to others is also time-consuming. Therefore, there is no orientation or pre-service programme for us to learn. Then we have to learn at the same time while doing. We steal too much time from our friends when we ask for help. Due to lack of time, we have to spend our personal time on extra work.

Constant system change (N=273) has the second highest frequency among the institutional challenges especially testing members face. Since testing units do not have a legal place within the organisation, such units can be affected by changes at the university, school, and unit levels. When the university administrators change, the school director and/or deputy directors can change. Thus, the decisions taken by the new leaders seem to change the testing procedures. In addition, the changes in the selection of books or the exam format can also influence testing practices. These changes were claimed to lead to instability in the system. A testing member expressed her concerns about constant systemic changes:

Her sene revizyona gidiyoruz zaten. Sürekli bir devinim var yani, öyle hani bu sistem oturdu artık, hiç oynamayalım gibi bir şeyimiz yok. Mesela kaynak sürekli değişiyor, yani her sene kullandığımız kitap değişiyor ve bu sebeple sistemimiz de değişiyor.

We make revisions every year. There is constant change, we don't have a settled system so that we don't need to make a change. For example, the source is constantly changing, that is, the book we use changes every year, and therefore our system also changes.

One of the testing members added further:

Son zamanlarda çok sık bir şeyler değişiyor. Ayşe'nin de dediği gibi iki yıl önce B2 düzeyinde olması gerektiğini düşündük ve final sınavı eklendi, ama şimdi sınavın bazı bölümlerini değiştirsek mi diye düşünüyoruz. Eylül'den sonra muhtemelen proficiency sınavda bir tekrar düzenleme yapılacak ve sitedeki örnek değişecek, öğrencilerin bilgilendirilmesi gerekecek. Her çeyrek komple değişiklikler olabiliyor, her çeyrek sıfırdan yazmalar veya modifikasyonlar gerekebiliyor. Çok sık yaşadığımız şey bunlar çünkü çok sık değişiklikler yaşıyoruz. Bu geniş çaplı değişiklikler bize problem yaratabiliyor. Daha biz bir örnek sınav çıkaramadan bir şeyler değişebiliyor.

There have been a lot of changes lately. As Ayşe said, we thought it should be at B2 level two years ago and then the final exam was added, but now we are thinking about changing some parts of the exam. After September, there will probably be a revision in the proficiency exam and the example on our website will change, students will need to be informed. There may be complete changes every quarter, thus we make rewrites or modifications. This is what we often experience because there is a change so often. These large-scale changes can create problems for us. Something can change before we can prepare a sample exam.

One testing member specified her concern on this issue as follows:

Her sene farklı muafiyet sınavı hazırlanıyor. İçinde listening, grammar, vocabulary ve tek bir parçalık reading var. Daha kısa bir sınav. Bir anda sınav bir tık daha uzun olsun deniliyor. Her sene bir değişiklik yapıyoruz. Mesela soru tipleri değişti, Proficiency'de iki aşamalıya geçtik, normalde note-taking yapmıyorduk, şimdi note-taking yapıyoruz. Bize söylenen ise sistem değişikliğine gittik. Geçen senenin sınav türüyle şimdiki farklı. Son beş seneye bakalım neredeyse benzerlik yok.

Each year, a different exemption exam is prepared. It included listening, grammar, vocabulary, and one reading text. A shorter exam. It is said that the exam should be one click longer. Every year, we make a change. For example, the question types have changed, we have changed to two stages in Proficiency, we didn't have note-taking, and now we are preparing a note-taking part. What we have been told is that we have changed the system. Last year's exam type is different now. If we compare the exams conducted last five years, there is almost no similarity.

As most testing members stated, poor physical conditions (N=254), including inappropriate working rooms, and crowds of classes make the process more difficult for testing members. During the interviews, it was frequently stated that poor conditions were having negative impacts on students' performance in exams. One testing member

was concerned about the room she is working in complaining about how small it is and that there is not enough space as it is full of books, old exam papers, and equipment. She explained her complaint:

Binamız bazen yetersiz kalabiliyor, bu kampüste ek bina yapıldı ama o ek bina başkalarına veriliyor sürekli, biz kullanamıyoruz. En son mesela, en fazla 25 kişinin olması gereken bir sınav salonunu 31 kişiyle yaptık. Bazı aylarda 25 bile fazla oluyor, güvenliğini sağlamak açısından veya iklim koşullarından. Örneğin, sıcaklığın 30'un, hatta 40'ın üstüne çıktığı dönemde 30-31 kişinin birkaç saat aynı ortamda kaldığını ve sınav çözdiğünü düşünün. Öğrencinin performansını bile etkiliyor. Ek bina yapıldı ama farklı bir kampüste. Dolayısıyla, kampüslerin uzak olması sorun oluşturuyor. İki ayrı kampüste sınav yaptığımızda, sınavlar hem kuzey kampüse hem güneye gitmek zorunda. Biz sınav günleri testing elemanları olarak yetişmekte güçlük çekiyoruz her iki kampüse. Üç haftamız var demiştik ya, diyelim ki sınavımız pazartesi günü; o gün bir aksilik çıkmasın diye diğer kampüse sınavları buradan cuma günü götürmek zorundayız. Bu yüzden perşembe günü her şeyin bitmiş olması gerekiyor. Pazartesi günkü sınav için perşembe akşamı her şey paketlenmiş, hazırlanmış, dosyaların yapılmış olması gerekiyor. Yani farklı kampüslere gitmek hem zaman daraltıyor hem de ekstra iş çıkartıyor maalesef.

Our building can sometimes be inadequate, an additional building was built on this campus, but that additional building was given to another department, and we cannot use it. In a recent example, we conducted an exam with 31 people, although the class size was a maximum of 25 people. In some months, 25 is even high, considering the security of the exam or climatic conditions. For instance, take into account that 30-31 people stay in the same environment for several hours and solve an exam when the temperature rises above 30 or even 40. This even affects the performance of students. An additional building was built but on a different campus. Thus, the remoteness of the campuses is a great problem. When we make an exam in two separate campuses, they have to be carried to both the north campus and the south one. We, as testing staff, have difficulty reaching both campuses on exam days. Let us say we have three weeks, and our exam is on Monday; we have to carry the exam papers from here to the other campus on Friday so that nothing goes wrong that day. Hence, Thursday everything should be finished. For an exam to be done on Monday, everything should be packed, prepared, and filed on Thursday evening. That is, different campuses both shorten our time and create extra work, unfortunately.

Some of the testing members indicated that anxiety appeared while attempting to work in an insufficient room:

Mekan olarak sığamıyoruz, çok eski, çok tozlu ve çok dar. Eskide kitaplıkmış, kütüphaneye dönüşmüş, böyle bir özelliği de var buranın, kapımızda da yazıyor. Buraya sığmıyoruz ve burada sınav hazırlaması oldukça zor. Böyle çok kalabalık bir sınav olduğunda ve arşive

gönderemediğimizde çünkü sıkıntı çıkmasın diye belgelerin bizde durması gerekiyormuş bir sene falan. O yüzden burada bekletmek zorundayız. Sürekli yeni sınavların yapıldığı, öğrenci sayısının fazla olduğu düşünüldüğünde fiziksel olarak burası bizim için çok küçük, çok eksi ve hiç kullanışlı bir oda değil. Mesela altı kişinin atandığını düşünün sınav birimine ama bu altı kişi nereye oturacak. Bu odada oturamazlar çünkü odanın içerisinde fotokopi makinesi, optik makinesi, artı sınavlar ve kitaplar duruyor. Bir de çalışma arkadaşın alerjisi varsa mesela bir arkadaşımızın toz alerjisi vardı. Hayatı cehennem oldu yani. Sürekli damlayla, antihistaminikle duruyordu. Bu oda bize yetmezken arkamızda gördüğünüz belgeleri koyacak yerimiz yok.

We can't fit into this place, it's too old, too dusty, and too narrow. It used to be a room for books, then it was turned into a library, and there is a tag on our door. We do not fit in here and it is very difficult to prepare an exam here. When there is such a big exam and we could not send the documents to the archive because we have to keep the documents for one year so that there would be no trouble. Therefore, we have to keep them here. Considering that new exams are being held and the number of students is high, this room is physically very small, very negative and not a useful room for us. For example, imagine that six people are assigned to the testing unit, but where will these six people sit? They can't sit in this room as there is a photocopy machine, an optical machine, plus exams and books in the room. What if your colleague has allergies? For example, one of our friends had dust allergies. Thus, his life was like hell. He was constantly using drops, of antihistamines. While this room is not enough for us, we have no place to put the documents you see behind us.

In another case, a testing member was anxious about technical problems in the building as she explained the situation in the interview as follows:

Eski çalıştığım kurumlarda farklı durumlar var. Bir üniversitede reading parçasını biz okuyorduk, diğerinde küçük hoparlörler kullandık. Bundan önceki üniversitede ise bilgisayar verildi ve flaş ile sınav yapıyorduk. Bu okulda ise binamız eski olduğundan teknik sorunlar yaşıyabiliyoruz zaman zaman. Mesela intercom sistemimiz var ama çok eski. Her sınavda bir heyecanla bekliyoruz, 'bir şey çıkacak mı, çıkmasın lütfen' diye. Hoparlörümüz yenilendi ve bilgisayarımız değişti ama tesisat eski. Geçenlerde kontrol etmiş olmamıza rağmen sınıfın birinde bir problem çıkıverdi mesela. Tabii intercomumuzun olması çok güzel bir şey ama bu tür şeyler yenilenmedikçe zaman içerisinde elektrik kesintileri oluyor. Binamız çok eski olduğundan tesisatta o kadar eskidir. 90'lı yıllardan falan kalma olabilir. Sınıftaki öğrenci sıralarının bile yarısı kırık dökük. Devlet okullarının birçoğunda böyle problemler vardır tabii ki. İşte tahtalar eski, projeksiyon varsa birçok sınıfta problem yaratıyor. Bu arada önceki yönetim ilgisiz ve kayıtsızdı. Ancak bu yönetim farkında ve mümkün olduğunca çözüm bulmaya çalışıyor ama elinden bir şey gelmiyor. Devlet okullarında yetersizlikler olabiliyor tahminimce.

There were different situations in the institutions I worked for in the past. At one university we were reading the reading text, at another one we used small speakers. At the previous university, we were using computers and conducting exams with flash. In this school, we may experience technical problems from time to time, since our building is old. For example, we have an intercom system, but it is very old. We are anxious in every exam, thinking "Will something come out? please not." Our speaker was renewed and our computer was changed but the plumbing is old. For example, although we checked it recently, a problem arose in one of the classrooms. Of course, having an intercom is a very good thing, but unless such things are renewed, there are electrical sections over time. Since our building is so old, the plumbing is old, too. It could be from '90s. Even half of the student desks in the classroom are broken. Of course, most public schools have such problems. Here, the boards are old. If there is a projection, it creates problems in many classrooms. Meanwhile, the previous management was disinterested and indifferent. However, this administration is aware and tries to find a solution as much as possible, but there is nothing they can do. I guess there may be inadequacies in public schools.

A number of participants indicated in the interviews that limited decision-making (N=226) seems to make them experience the same challenges within a vicious cycle. In general, school administrators, especially in state universities, complained about the limited representation power of the language schools even while taking a decision about those schools. One director mentioned in the interview:

Yabancı diller okulunun temsil gücü maalesef üniversitelerde akademik takvimden tutun da diğer şeylere kadar hep böyle ihmal edilen boyutta oluyor. Sürekli eğitim merkezlerinden bile daha geride oluyor. Çünkü düz mantıkta SEM'ler para getiriyor. Bizim okulumuzun senatoda da söz gücünün zayıf olması sebebiyle problemleri biz yaşıyoruz ve çözmek zorunda kalıyoruz ama elimiz kolumuz bağlı kalabiliyor. Akademik takvim bile bizim sisteme uymayabiliyor.

Unfortunately, the representation power of the foreign language school is generally neglected in universities, from the academic calendar to other things. It is even behind the continuing education centres. That's why SEMs bring money. Due to the weak voice of our school in the Senate, we experience problems and have to solve them on our own, but we are in a desperate situation. Even the academic calendar may not suit our system.

Another point was mentioned by another director:

YADYOK'a danışılmadan alınan karar neticesinde mesela YADYOK'u buradan alçak katlı binaya taşıdık. Bize hiç haber verilmeden alındı karar. Bize sadece şey dendi 'Eşyalarınızı toplayın kuzeye gidiyorsunuz.' En yoğun olduğumuz Eylülde geldiğimizde burada güya bina tamir ediliyordu. Bir de demokratik bir üniversite olduğumuz söyleniyor. Oldu bittiye

getirilerek tek katlı bir binaya taşındık. Biz defalarca gidip o binada hazırlık okulu eğitimi verilemeyeceğini anlatmaya çalıştık rektörlük düzeyinde. Toplantılar yapıldı defalarca ama anlatamadık. En acısı da öğrenci tamamen izole, üniversite hayatından tamamen kopuk bir alana gönderiliyor.

Following the decision taken without consulting YADYOK, for example, we moved YADYOK from here to a low-rise building. The decision was taken without informing us. We were only told, 'Pack your things, you are going to the north.' In September, when we were very busy, the building was supposedly being repaired here. It is said that we are so-called a democratic university. We moved to a single-floor building, which brought us to a dead end. We have tried to explain at the rectorate level that preparatory school education cannot be given in that building. Meetings were held many times, but we could not explain. Worst of all, the students are sent to a completely isolated, completely disconnected area from university life.

Another director added in these words:

Biz iki sene önce finalin ağırlığının düşürülmesini önermiştik ama o kadar basit bir karar olmadığı bize söylendi. Bu sene ise öğrenci sayısı çok artınca, rektörlükten gelen baskıyla bir komite oluşturulmuş ama YADYO'dan kimse yok bu komitede, eğitim fakültesinden akademisyenler var. Eğitim fakültesindeki kişilerin çoğu da haberdar değil aslında YADYO'nun iç süreçlerinden. Testingden ya da YADYO'dan hiç kimseye sormadan o komite kendi kafasından karar veriyor. Onayı zaten yönetim, yani Rektörlük veriyor.

Two years ago, we suggested reducing the weight of the final, but we were told it was not such a simple decision. This year, when the number of students increased a lot, a committee was formed under pressure from the rectorate, but there was no one from YADYO in this committee, but academicians from the Faculty of Education. Most of the people in the Faculty of Education are not aware of the internal processes of YADYO. That committee decided on its own, without asking anyone from testing or YADYO. The approval was already given by the administration, namely the Rectorate.

Another major concern of the participants within institutional challenges is the problems with the necessary equipment (N=205). In some universities, the computers used to prepare questions are limited, thus the fact that testing members use them one by one can set back their schedule. For instance, a testing member working at a state university emphasized that he had to go to the campus at the other end of the city in his own car in order to make photocopies of the exam. However, if they had such a machine, they could spare this time for their work. He explained his stress about the lack of equipment in the interview:

Genellikle devlet üniversitelerinde yaygındır, özellerde de yönetim desteği yoksa bu sorunlar yaşanmaktadır. Sınav sürecimizde bilgisayarlar, fotokopi makineleri, optik okuma cihazları, flaşlar veya veri saklamak için harici bellekler gibi ekipmanların önemi o kadar fazla ki olmadığında çaresiz kalabiliyoruz veya kendi imkanlarımızla çözmeye, telafi etmeye çalışıyoruz. Bazen elimizdeki makinelerin de bozulmaması için dua ediyoruz. Bir servis gelecek de ilgilenecek ama ne zaman? Mesela geçenlerde harici bellek alalım dedik, ama verilmedi, sınavları yedekleyemedik. Bilgisayarımız çöktü, sonrasında bir sürü veri kaybettik.

Problems with equipment are generally common in public and private universities if there is no management support. In our exam process, Equipment such as computers, photocopy machines, optical reading devices, flash drives, or external data storage drives is so important that if we lack, we can be helpless or try to solve it with our own means. Sometimes, we pray that the machines in our hands do not break down. Then a service will come and fix it but when? For example, recently we wanted to buy an external memory, but we were not given one, so we could not back up the exams. Our computer crashed, and we lost a lot of data.

Furthermore, limited financial support (N=179) is a noticeable concern of the participants. Lack of necessary financial contribution for technology-based applications and other expenses that can be used for office needs and exam processes seems to demoralize few employees working in these units and make their work more difficult. Having a lack of financial or emotional external contribution, TUMs are left to their intrinsic motivation to fulfil their responsibilities. A testing member stating that there were paid applications to find resources or to be used in statistical work, but they could not afford them elaborated on the issue as follows:

Maddi anlamda pek testinge destek olunmuyor aslında. Alt düzeyde yani işin mutfağında ihtiyaç duyulan bir şeyin görmezden gelinmesini anlamıyorum. Alttan bir talep gidiyorsa belli ki bir ihtiyaç vardır. Ama maalesef böyle yürümüyor işler. Böyle bir şey oluyor maalesef. Bir sürü şeye para harcıyoruz ama ek ödenek dediğimizde yok deniliyor. Alınması gerekenleri listeliyoruz, ancak kaynak yetersiz deniliyor. Bu ofiste neden çalışalım o zaman. İkna yoluyla, zorla veya kendi iç motivasyonunuzla bir şeyler yapmak zorunda kalıyoruz. Teşvik çıkarılıyor ancak biz bu yoğunlukta çalışma yapamıyoruz ve faydalanamıyoruz. Bu görevlerin bir şekilde karşılığının olmaması, üstüne ihtiyaç duyduğumuz şeylerin alınmaması çok saçma.

In fact, there is not much financial support for testing. I don't understand how they ignore a need at a lower level, namely in the kitchen of the school. If there is a demand from the bottom, obviously there is a need. But unfortunately, things don't work that way. We spend money on a lot of things, but when we ask for additional allowance, we are told there is not.

We list the must-buys, but the resource is said to be insufficient. Why should we work in this office then? We are compelled to do something by persuasion, by force, or by our own intrinsic motivation. Incentives are provided, but we cannot conduct studies under these conditions, thus we cannot benefit from them. It is absurd that these duties do not have a reward in some way, and that the things we need are not given.

Another point made by the participants was their concern about insufficient coordination (N=157) among the units within the school. In addition to testing units, there are other units such as curriculum units, learning management units, and whatsoever. However, unless these units share their work with other units or carry out joint work, the examination processes are adversely affected. In addition to the workload in the exam unit, extra work may arise due to a lack of coordination. It was especially emphasized in the interviews to establish good coordination among the units on such issues as the preparation of test specs and the determination of the exam contents. A testing member, for instance, reported:

Testing ofisi olarak biz curriculum ile bağımlı çalışmalyız. Ancak bu birimlerin yasal yapıları olmadığından ve çalışan kişiler sürekli değiştiğinden birimler arasında kopukluklar oluyor. Onlar bir şey hazırlıyor ama bize danışmadıkları için yanlış veya hatalı olabiliyor. Bize haber vermeleri gereken durumları es geçebiliyorlar, sonrasında haber verildiğinde ise çok geç olabiliyor. Mesela geçenlerde konu bakımından bir yere kadar öngörüp sınavı hazırladık. Birden curriculumdan bir cevap geldi, bazı sınıflar oraya kadar gelememiş.

As a testing office, we should work with the curriculum. However, since these units do not have a legal structure and the people working at the units are constantly changing, there is insufficient coordination between the units. They prepare something, but because they do not consult with us, it may be wrong or inaccurate. They may disregard informing us of some conditions. When we are informed later, it may be too late. For example, we recently prepared an exam predicting the subject to a certain extent. All of a sudden, there was a reminder from the curriculum that some classes could not get there.

In addition to the challenges stated above, an incompatible automation system (N=132) is a factor contributing to breaking the order of the testing units. Grading systems are generally processed within an automation system developed by a company through outsourcing in most universities. However, the automation systems developed by these companies, upon which universities have agreed, seem to fall short in calculating the scores of the preparatory programmes such as quizzes, midterms, and

out-of-class or in-class activities. Although this situation is constantly expressed by testing members at preparatory schools, their demands are responded negatively because the companies ask for extra fees or state that they will have to provide additional services. As a result, score calculations are entered manually or announced in a different way rather than automation. One of the testing members elaborated on the existing automation system and stated:

Hazırlık okulu olarak Proliz sistemi bize hiç uygun değil. Çünkü 1000 kişiye sınav hazırlıyoruz ve 1000 kişilik sınav yapıyoruz. Derse giren hocaların tek tek açıkladığı bir sistem bize uymuyor. Bir de bizim quizlerimiz var, quizlerin farklı oranları var. Örneğin, bir reading quizinde tasklarımız var, bu taskların da yüzde 25'i çarpılacak ama Proliz sistemi onu yapamıyor. Yapamadığı için biz de bütün notları sistemden çekip elden hesaplıyoruz. Bu da bizi uğraştırıyor çünkü reading quizi yüzde beş ama yüzde beşin de kendi içinde bölümleri var. Hazırlıklar için uygun bir sistem bulamıyoruz çünkü piyasadaki en ucuz sistem bu olduğu için okul bunu kullanıyor. Yönetim de bizi dikkate almıyor.

As a preparatory school, the Proliz system is not suitable for us at all because we are preparing and conducting exams for 1000 people. A system in which instructors can announce the scores of their classes does not suit us. We have quizzes, which have different weights. For example, we have tasks in a reading quiz as 25 percent of these tasks should be calculated, but the Proliz system cannot do it. Since it cannot, we download all the scores from the system and calculate them by hand. This bothers us because the reading quiz is five percent, but there are subsections in that five percent. We cannot find a suitable system for preparatory schools because it is the cheapest system on the market so our school uses it. Rectorate does not take us into account.

A director explained the issue in a more detailed way:

Teknik olarak büyük bir problem aslında mesela OBS sistemine biz 70 sınıfla başlıyoruz. Hazırlıkta 70 tane sınıfın tek tek notlarını giriyorum. Hop diye okulun notlarını açıklayamıyoruz. 70 sınıf olunca bir sınıfı unutuyorum, geri dönmesi zor oluyor, karmakarışık bir sistem oluyor. Çok büyük bir problem, notların OBS'de aynı anda açıklanması için hocalara sürekli duyuru yapıyoruz, ben tek tek notları e-mail atıyorum hocalara, diyorum ki ben size notları bugün gönderiyorum ama çarşamba günü aynı anda açıklayalım. Bakıyorum bazı hocalar salı günü açıklıyor. Genellikle bu sistem çoğu üniversitede kullanılıyor. Ama bu sistem hazırlığa uymuyor bence. Biz hoca bazında sınav duyurusu yapmıyoruz, testingde 1000 kişiye sınav yapıyoruz. 70 tane sınıfı tek tek girmem gerekiyor ama zaman probleminden ben de kontrol edemiyorum yani.

Technically, it is a big challenge, for example, we enter 70 classes into the OBS system. In our preparatory school, I enter the scores of 70 classes one by one. We cannot announce the scores automatically. When there are 70 classes, I may forget a class, then it is difficult to

return, so it becomes a chaotic system. It is a very big problem that we are constantly making announcements to the instructors so that the scores will be announced on OBS at the same time, I e-mail the scores one by one to the instructors and I firmly tell them that I am sending the scores this day, but we will announce them at the same time on Wednesday. I realise some teachers are announcing on Tuesday. Generally, this system is used in most universities. But I think this system is not suitable for preparatory schools. We do not announce exam scores on the basis of individual instructors, we prepare exams for 1000 students in the testing office. I have to enter 70 classes into the system one by one on my own, but I cannot check them because of the time constraint.

The circle of the same implementations along with the same mistakes is believed to centre on the lack of a feedback system (N=107) from instructors and students. Acknowledging the benefits of a well-developed feedback system, testing members approached the issue from the perspective of the need for change. They seem to feel desperate and demotivated to be stuck in time constraints and heavy workloads, which they believe could be conveyed to rectorates through feedback systems so that something can be changed. When there is no formal feedback system, informal feedbacks appear to be mostly negative clashes between TUMs and instructors or students. A testing member, for example, considered:

Resmi bir dönüt sistemimiz yok. Bireysel olarak hocalar, bazen de öğrenciler testing ofise gelip şikayetlerini bildirebiliyorlar. Öyle bir rahatlıkları var. Genelde de memnun olmadıkları şeyleri veya hoşlarına gitmeyen konuları bireysel olarak bildiriyorlar. Daha kötüsü zaman zaman idarecilere şikayet ediyorlar. Fakat gerekli değişikliklere ışık tutabilecek bir anket sistemimiz yok.

We do not have an official feedback system. Individually, instructors and sometimes students can come to the testing office and make their complaints. They have such comfort. In general, they mention things that they are not satisfied with or that they do not like individually. What is worse, they may complain to the administrators from time to time. But we do not have a survey system that can shed light on the necessary changes.

One of the testing members added that:

Biz öğrencilerden öğretmenler yoluyla veya e-kışisözlük sayesinde dönüt almış oluyoruz. Genelde zorlandıkları kısımlar veya süre ile ilgili şikayetleri oluyor. Mesela bir sınavda grammer bölümündeki sorunun çok zor olduğunu, öğrencilerin yapmakta çok zorlandığını ve sürenin çok az geldiğini bildirdiler. Biz de sonuçları istedik hocalardan hani o bölüme dair öğrenciler kaç puan almışlar bakalım diye. Sonuçları geçmiş yılların notlarıyla

karşılaştırdık. Değişiklik yapmamız gerektiğini gördük. Aslında düzgün bir anket sistemi olmadığından boş iddialar veya şikayetler oluyor.

We receive feedback from students through teachers or the “ekşisözlük.” In general, students have complaints about the parts they have difficulty in or the duration. For example, they reported that a grammar section of an exam was very difficult, the students had a hard time doing it and the time was short. We asked the teachers for the results, to see how many points they got for that part. We compared the results with the scores of previous years. We saw that we did not need to make a change. In fact, there are pointless claims or complaints when there is no proper survey system.

One of the testing members emphasized:

Öğrenci anketlerimiz mevcut, bir de arkadaşlarımızın birbirini değerlendirdiği anketler mevcut. Fakat bu anketlerde ve değerlendirmelerde testing birimi için doğrudan bir kısım yok ve dolayısıyla sonuçlarla ilgili bizim ofise verilen direkt bir bilgi olmuyor.

We have student surveys, and there are also evaluations through which our friends evaluate one another. However, in these surveys and evaluations, there is no direct part for the testing unit, so there is no direct information given to our office about the results.

Some of the testing members approached the challenges by emphasizing the lack of evaluation systems (N=64), which directs criticisms to tests and TUMs. According to them, instructors give their lectures for a semester or a year, yet it is not evaluated whether they used class time effectively or if they are telling unnecessary things. It is not evaluated whether they explain the subjects well enough. Only if a student specifically and personally has problems and complains, the relevant teacher is evaluated. In fact, exams also reflect the performance of the instructors to some extent. In addition, the exam results are perceived only as “passed or failed; easy or difficult.” However, exams are a tool for improving the process. As stated by one of the directors:

Önemli sorunlardan biri de sınav sonrası evaluation boyutu. Bence ciddi challengelardan biri hocalar açısından, sistem açısından after the administration of the exam the evaluation part is challenging. Bu sürecin Türkiye’de de biraz ihmal edildiği düşünüyorum. Herkes sadece sınava odaklanıyor. Sınav bitiyor, sınav zordu, sorular yanlış algısı oluşturuluyor. Hocalar değerlendirilmiyor maalesef. Sınav sonrası ile ilgili değerlendirmelere istenilen kalitede zaman ayrılmıyor.

One of the important problems is the evaluation dimension following the exam. I think this is one of the serious challenges from the point of view of the instructors and from the point of view of the system. After the administration of the exam, the evaluation part is challenging. I think that this process has been neglected in Türkiye. Everyone just focuses

on the exam. Once the exam ends, the perception is disseminated as the exam was difficult, and the questions were wrong. Unfortunately, instructors are not evaluated. The required quality time is not allocated for post-exam evaluations.

One of the testing members focused on:

TOEFL'da her çıkan seviyenin karşılığı bir descriptive var. Hani bizde böyle bir şey yok. Biz sadece geçti, kaldı veya B1, A2 diyoruz. Keşke öğrencilere, hocalara sınav sonuçlarına ilişkin nitel açıdan ve nicel açıdan anlamlı çıktılar sunabilsek.

There is a descriptive part corresponding to each part in TOEFL. We do not have anything like that. We just announce as passed, failed or B1, A2. I wish we could provide our students and instructors with qualitative and quantitative meaningful outcomes regarding the exam results.

There are some other participants who expressed views on the challenges of the term-based system (N=43). While acknowledging the advantages of the modular systems, most of the state universities and several foundation universities employ term-based or year-based systems due to a number of reasons. Hence, the students having language education in the term-based system may remain in the same class with their classmates regardless of their interests, pace, progress, and level. School directors mostly complained that both fast learners who are bored and slow learners who do not keep up with the pace do not belong to the class and tend to have private courses for central exams. The following extract exemplifies how the system affects the practices:

Term based sistem kullanmamızdan dolayı seviyeler arası geçişte problem yaşıyoruz. Öğrencilerden, hocalardan ve okulumuz bünyesinde oluşturduğumuz birimlerden sistem ile şikayetler alıyoruz. Ancak moduler sisteme geçemiyoruz. Hoca sayımız yeterli değil. Az sayıda hoca ile ancak term based kullanıyoruz. Mesela geçenlerde curriculum birimi hocalar çok yüklenmiş, şikayet ediyorlar yetiştirememekten. Hocalar yetiştiremedikleri için haliyle geride kalan öğrencilerle de ilgilenemiyorlar.

Due to using the term-based system, we encounter problems in transitioning between levels. We receive complaints against the system from students, instructors, and the units we established within our school. However, we cannot switch to the modular system. We do not have enough teachers for that. We can only use term-based with few instructors. For example, the curriculum unit has recently delivered overload to the instructors who are complaining about not being able to keep up with the curriculum. Since they cannot keep up, the instructors cannot deal with the students who are left behind.

A testing member further elaborated on the specifics of the term-based system and explained in detail:

Bizim sistemde öğrenciler aynı seviyeyle devam ediyorlar dönem sonuna kadar. Yani adları aynı kalıyor, kendileri ilerliyorlar. P2 başlıyor işte P3'ün sonuna gelmiş oluyor falan. Ama adı P2 olarak kalıyor, değişmiyor, yani seviyeler arası geçiş yok. A1'ler A2 ve A2'ler aslında gerçekten B1 olmuyorlar. İkinci dönem A1 ve A2 seviyeleri B1'e geçiyor. Aynı kitabı okumuş oluyorlar, ama tabii sonuçta öğrencinin seviyesi, başlangıçları bir değil. Başka bir deyişle, sene başında bir grup A1, bir grup da A2 oluyor, sonra bahar döneminde aynı seviyede oluyorlar. B1'den B2'ye geçenlerde hiçbir problem olmuyor ancak özellikle B1 öğrencilerinde sistemsal açıdan çok problem yaşıyoruz. İşte A1'den B1'e atlayınca bazı öğrenciler ilerde, bazıları da geride kalıyor. İlerde olanlar sıkılıyor, geride kalanlar çok eksik olduğundan onlar için ağır bir program oluyor.

In our system, students continue with the same level until the end of the semester. That is, their names remain the same although they move forward. They start as P2 and pass into P3 in the end. But its name remains P2, it does not change, so there is no transition between levels. A1s are not A2 and A2s are not really B1. In the second semester, A1 and A2 levels move to B1. They read the same book, but of course, the student's level and beginnings are not the same. In other words, at the beginning of the year, one group becomes A1 and another group becomes A2, then they are at the same level in the spring term. There is no problem for those who pass from B1 to B2, but we have a lot of systematic problems, especially with B1 students. Here, when jumping from A1 to B1, some students stay ahead and some lag behind. Those who are ahead are bored; on the contrary, those who are behind are weak, so it becomes a heavy schedule for them.

Finally, some of the participants pointed to the issue of limited assessment of transferable skills (N=26). They perceive that English is taught and tested through rules even at the university level, but language skills within the 21st-century skills cannot be transferred to communication and daily life, therefore remain abstract, which they believe seems to result from rarely assessing transferable language skills. A testing member put an emphasis on limited assessment of transferable skills and stated:

Tıpkı eğitimimizde olduğu gibi hazırladığımız sınavların da birçoğu maalesef transferable becerileri ölçmüyor veya çok sınırlı. Öğrettiğimiz gramer kurallarını sorguluyoruz, ezberlediler mi acaba? Sık sık 21. yüzyıl becerileri deniliyor ama bizim sisteme pek uygun değil mi acaba?

Just like in our teaching, many of the exams we prepare unfortunately do not test transferable skills or do so in a very limited way. We test the grammar rules we have taught, have they memorized them? It's frequently mentioned 21st-century skills, but isn't it well suited to our system?

Out of 12 institutional challenges, pre-service or in-service training was the top issue. Due to several reasons from financial lack to insufficient number of experts in language testing, the directors of EPPs were complaining about a huge gap between ideal standards and the real ones. Besides, most TUMs criticised system changes with the change of the leaders such as the rector, the vice rector or the director. Another major issue was about lack of necessary materials, human power, financial contribution. Due to these major challenges, TUMs were seen to be desperate. In the next section, unit-related challenges will be explained.

4.1.1.3. Unit-related challenges

This subcategory includes nine emerging codes which are shown in Table 4.5.

Table 4.5. *Challenges at the unit level*

Emerging Codes	Number of Codes	%
Excessive workload	359	19,72
Being understaffed	311	17,09
Excessive TUM turnover	263	14,45
Need for support staff	230	12,64
Lack of orientation	202	11,10
Ambiguity in the operational system	176	9,67
Need for a comfortable work environment	124	6,81
Ineffective coordination among TUMs	89	4,89
Problems with the previous system	66	3,63
Total	1.820	100

The challenges related to excessive workload (N=359) comprise the most frequently used type in this subcategory in the interviews. Apart from the work in preparing exams, there are a number of tasks in the organisation and post-implementation process that are shared among limited members who also teach in the classroom. TUMs mostly complained to have been carrying the most difficult part of the process namely item writing to their home due to other work which should be done only in the office. In addition, they also mentioned that they could not have time to have lunch and instead consumed snacks or biscuits so as to save time for meetings. The following extracts from testing members' experience about their stress when the term starts would be enough to clarify the issue:

O kadar yoğunuz ki genelde zaman yetişmiyor. Özellikle dönem sonlarına doğru sınavlar üst üste binmiş oluyor. Aslında iyi kötü bir programımız var ama o kadar yoğun bir iş yükü var ki işler üst üste binince yetiştiremiyoruz. En nihayetinde yetiştiriyoruz ama özel hayatımızdan, okuldan sonraki kalan vaktimizden feragat ederek yapıyoruz bunu. Görüyorsunuz sizinle mesela şu an oturup görüşürken bile şunu (kağıtların ayrılması) da yapmak zorunda kalıyoruz. Genellikle durum böyle oluyor. Bizim şey şansımız hiçbir zaman olmuyor, burada ofis olarak hadi gelin bir hava alalım, şurada bir kahve içelim. Çoğu zaman yemek yemeye fırsatımız bile olmuyor, burada atıştırma şeklinde oluyor. İnanın yemekhaneye gittiğimiz günler sayılı çünkü zaman kaybı. Yemekhaneye gitmek bizim için bir saat demek. Gideceksiniz yemeği yiyip buraya tekrar döneceksiniz. Yılda toplam kaç defa gitmişizdir on kere gitmemişizdir. Yeri geliyor, yoğunluktan kendimizi kaptırıp yemek yemiyoruz.

We are so busy that we are often out of time. Especially during the end of the semester, exams overlap. In fact, we have more or less a good schedule, but there is such a heavy workload that we cannot keep up when work gets busy. In the end, we complete, but we achieve this by sacrificing our private life, and the rest of our time after school. As you see, for example, we have to do this (separating the papers) even when we are sitting and talking with you right now. That is usually the case. We never have a chance, as an office, to get some air and have a coffee somewhere. Most of the time, we do not even have the opportunity to eat, but for snacks. Believe me, the days we go to the cafeteria are limited because we see it as a waste of time. Going to the cafeteria means one hour for us. We will go, eat food and come back here again. We have not gone ten times this year. We are so stuck with the workload that we forget to eat.

Yoğunluğumuz anlatmakla bitmez. Sınavları biz basıyoruz, biz paketliyoruz. Şurada bekliyor beni hatta şu an, bunlar zarflanacak. Sınavları hazırlıyoruz. Sınav hazırlaması bittikten sonra atıyorum fotokopi makinesi bir sayfayı yanlış çekti, bunun kontrolünü kim yapıyor? Makinenin toneri bitmiştir, kayma olmuştur, boş çıkmıştır. Bir sayfası boş çıkarsa biz uğraşıyoruz. Zarfa giren bütün materyalden biz sorumluyuz. Şurada speakingler basılmış şimdi onlar zarflanacak hepsi benim sorumluluğumda mesela. Sınavların hazırlık aşamasındaki görüşmeleri, toplantıları, proofları biz yönlendiriyoruz, gelen revizyonları yapıyoruz, basımı yine biz yapıyoruz, bastık zarflamasını yine biz yapıyoruz, hocaya biz veriyoruz, sınav koordinasyonunu biz yapıyoruz, itiraz gelirse ki geliyor ilgili kağıtları araştırıyoruz, buluyoruz, tekrar inceliyoruz, biz değerlendiriyoruz. Mesela itiraz writing de geldi, zaten iki hoca önceden okumuş oluyor, biz de bir okuyoruz. Sonra da sonucumuzu basıp öğrenci işlerine götürüyoruz. Raporunu biz veriyoruz yani. Bunlar aklıma geldiğince sürecin en yalın hali.

Our workload cannot be told. We print the exams, and pack them. For instance, they are waiting for me over there, even right now, these will be enveloped. We are preparing the

exams. After the exam preparation is over if the photocopy machine copies a page wrong, who checks it? The toner of the machine may run out, there may be an error of a shift or an empty page. If a page is blank, we deal with it. We are responsible for all material that goes into the envelope. For example, the speaking exams are printed here, now they will be enveloped, all of which are under my responsibility. We organise the interviews, meetings, and proofs during the preparation phase of the exams, we make the revisions, we print it again, after printing, we give it to the teacher, we coordinate the exam and if there is an objection, we search, find, review and evaluate the relevant papers. For example, the objection comes in writing, already two teachers have read it, but we are making the third reading. Then we print the results and send them to student affairs. That is we report it. These are the simplest process I can think of.

A testing member added that there is no work distribution as test writer and test administration at their unit and explained:

Testing birimi, yasal bir zemini olmadığından ve sınav komisyonu gibi çalıştığından bizim için sanki ekstra boş zamanlarımızda, zorunlu olarak yaptığımız bir şey gibi oluyor. Ders saatinde altı saatlik bir reduction yapılıyor yani diğer hocalar 18 saat derse girerken biz 12 saat derse giriyoruz. Ama onun dışında onların sorumlu olduğu her şeyden biz de sorumluyuz. Derse de giriyoruz, sınav gözetmenliğine de giriyoruz, sınıfımızın sınav kağıtlarını okuyoruz, portfolyolarımızı da takip ediyoruz, her şeyi yapıyoruz yani. Bu arada ofise 30 km uzaklıktaki A... 'a her hafta derse gidiyorum, gözetmenlik yapmaya gidiyorum, 40 kişilik sınıftan writing portfolyom geliyor ve 40 tane essay okuyorum. Sınav komisyonu bunların üstüne karşılıksız çek gibi verilmiş zorunlu bir görevlendirme.

Since the testing unit does not have a legal basis and remains like an exam commission, it is something we are required to do in our extra time. We are given a 6-hour reduction, that is, we teach 12 hours while other instructors teach 18 hours. However, we are also responsible for everything they do. We attend the class, do invigilation in the exams, read exam papers of our classes, follow our portfolios, I mean we do everything. For instance, every week, I go to A..., which is 30 km away from the office, to teach, and do invigilation in the exams there. Besides, I am sent writing portfolios from 40 students and I read 40 essays. The exam commission is a compulsory assignment given like an unpaid cheque.

The voice of a testing member also emphasized:

Bizim birimde test admin-test writer gibi bir ayrım yok. Biz dört kişiyiz, her şeyi yapıyoruz yani. Bir leveldan ben sorumluyum, sorusunu da ben yazıyorum, fotokopisini de ben çekiyorum, sınav görevlendirmesini de ben yapıyorum. Öyle olunca sadece sınav yazmaya pek odaklanamıyoruz. Aslında sınav hazırlaması, gerçekten çok zor ve dikkat gerektiren bir

iş. Ofiste hep yoğunuz yani. Kafamız hep yoğun olduğundan kendimizi sınava veremiyoruz. Onunla ilgili bir çözüm getirirseniz şahane olur.

There is no such distinction as test admin-test writer in our unit. We are four people, who do everything. If I am responsible for a level, I write questions, take a photocopy, and make exam assignments. In this way, we can hardly focus on writing test items. In fact, preparing an exam is a very difficult and demanding task. Yet we are always busy at the office. Because our minds are always busy, we cannot focus on exams. It would be great if you could come up with a solution for it.

Being understaffed (N=311) has the second highest frequency among the challenges in this subcategory testing members face when codes are computed. Considering the difficulty of language testing process and high tempo, the number of experts working in the offices is insufficient. Limited TUMs need to deal with overtime work, time constraints, and stress. All these are believed to lower their motivation. A testing member explained as follows:

Bizim en çok çektiğimiz sıkıntı açıkçası sınav hazırlama sürecinde burada sadece soru item yazmak değil de fiziksel olarak sınavı hazırlama sürecinde oluyor. Zaman sıkıntımız var ve tempomuz yüksek ancak dört kişi çalışıyoruz. Çok yoğunuz bu defa fazla mesai yapmamız gerekiyor hatta çoğu zaman. Geçen sene altıydık bu yıl dört olduk, zaman darlığı daha da arttı. Dört kişi burası için çok yetersiz, X... Üniversitesi çok kalabalık bir üniversite. Burada bir sınav birimi görevlisi başına 1000 öğrenciden fazla öğrenci düşüyor.

In fact, the biggest problem we face is not only writing an item, but also heavy physical work during the exam preparation process. We have time constraints and our tempo is high, but we are only four people. We are very busy, and thus we have to work overtime, even most of the time. Last year, we were six, but we became four this year, so time constraint has got more. The number of four people is not enough for this place, X... University is a very crowded university. There are more than 1000 students per exam unit officer here.

A testing member shared similar views but also put emphasis on permissions:

Sadece tek kişi bir seviyeden sorumluyuz ve hepimiz insanız. Geçen sene N... hocamızın bir ameliyat olması gerekmişti, o zamanlar A seviyesinden sorumluydu. Düşünebiliyor musunuz A seviyesinde normalde 2000'e yakın belki öğrenci var. Bu olay olduğunda net sayı 1700'dü. Şimdi tabii arada böyle bir dengesizlikte var seviyeler arasında. Öğrenci sayısı seviyeler bazında değişiyor, giriş seviyeleri daha kalabalık oluyor. Bir seviyeden sadece bir kişi sorumlu olduğunda herkes çok meşgul oluyor. Ama bu durumda kendi görevlerimize ek izinli olan kişinin sınavını da hazırlamak zorunda kalıyoruz. Herkes sağlık problemi yaşayabilir, ailevi bir problem yaşayabilir.

Only one person is responsible for one level, and we are all human. Last year, N... had to have an operation, she was responsible for the A-level at that time. Can you imagine there are approximately 2000 students at A level in general? The number was 1700 when that happened. Of course, there is such an imbalance between levels. As the number of students varies by level, entry levels are getting more crowded. Everyone gets very busy when only one person is responsible for a level. But in this case, we have to prepare the exam of the person who is on leave in addition to our duties. Anyone can have health problems, or a family issue.

One of the directors drew the attention to lack of new staff:

Rektörümüze de anlattım bu birimin minimum sekiz kişi olması lazım, bu birim dört kişi ile kesinlikle yürümüyor. Mesela şimdi öğretim görevlisi alımı yapılacak, çeviri bürosuna verilecekti. “Kaç kişi alıyorsunuz” dedim birmiş, beş kişi istenmiş, bir kadro verilmiş. 5-6 hocamız emekli oldu, ancak bir kişi gelebiliyor, o da bizim okula değil.

I also explained to our Rector that this unit should have a minimum of eight people, and this unit does not move forward with four people. For example, now a instructor will be recruited and given to the translation office. “How many people will you recruit?” I asked. We were given one though we asked for five people. 5-6 instructors have retired, but only one person can come, and that one not to our school.

The third most frequent anxiety in this subcategory was testing members' concerns about excessive testing member turnover (N=263). The interviews showed that the challenges such as the difficult and stressful examination process, the lack of additional income, and no systematic structure cause most of the TUMs want to leave the office after a few years. The worse part for directors might dwell on the seek to find new staff as no other instructors would be willing to agree to work in the office. Therefore, the assignment is maintained by force without even asking for their wills and there appears frequent TUM circulation in the office. A director elaborated:

Sınav hazırlamanın kendisi stres dolu, zor bir süreç ve artı bir ücreti de yok, dolayısıyla bu göreve genelde herkes gönüllü ve istekli olmuyor. Herkesin gözü testing member ın üstünde, O yüzden genelde grubun zaten %70 i istemiyor o işi yapmak. Genelde özellikle devlet olmak üzere bütün üniversitelerde böyle. Çok istenilen, tercih edilen yerler olmuyor. Çok büyük bir sorun sürdürülebilirlik. Bir kişi zorunlu iki yıl yapınca devam etmek istemiyor. Bu durumda testing ofisin sirkülasyonu çok fazla, yeni katılan insanlar da sürekli değişeceği günü bekliyor, askerlikteki gün sayma gibi.

Exam preparation itself is a stressful, difficult process and there is no additional payment, so not everyone is willing to work. Everyone's eyes are on the testing member, so usually

70% of the group doesn't want to do that job. In general, this is the case in all universities, especially in the state ones. These offices are not desired and preferred places. Sustainability is a huge issue. A person does not want to continue after having completed compulsory duty for two years. In this case, circulation in the testing office is frequent, and the newly recruited people are constantly waiting for the day to change, like counting the remaining days in the military.

Another director drew the attention to another side of the issue:

Rektör hocayla görüştüm, rektör hocayı da ikna ettim. Testing birimine iki kişi aldım, bunları yurtiçinde seminerlere gönderdim, döndüler özel bir üniversiteye eğitim için gönderdim, İngiltere'ye gönderdim iki haftalık eğitim aldılar. Ayrıca, yurt içindeki her konferansa istisnasız birer birer gönderdim bir yıl boyunca. Tam bunlardan verim alacağım, bir yıl sonra hocam ben ayrılmak istiyorum dediler. Şimdi tamam onlar da haklı, yoğun tempoda çalışıyorlar. Düşündüm bunları çıkartsam yerine bu sefer iki kişi alacam. Rektöre gidip ne diyeceğim? Gittim rektöre söyledim, rektör “biz onları eğittik ya, onlar devam etsin” dedi. Bu sefer de emir gibi oldu, çalıştılar ama gönüllü olmayınca da ne kadar sınav kaliteli çıktı sorgulanır. Sonra rektöre söylemeden kişileri değiştirdim, bu defa yeni işe başlayan kişileri aldık ve eğitime gönderemedik. Bazı sorunlar burayı aşıyor, YÖK'ün bunlardan haberi yok. Birilerinin bunları bir yerde uygun bir dille dile getirmesi gerekiyor.

I explained the issue to the rector and convinced him. I assigned two people to the testing unit, sent them to several seminars in the country, then to a private university for training, then to England, where they received two weeks of training. In addition, I sent them one by one to every conference in the country, without exception, for a year. When it was time to benefit from them, one year later, they told me to leave. In fact, they are right about working at a high tempo. I thought if I removed them, I would take two new people this time but what could I say to the rector? I went and told the rector, yet the rector said “we trained them, so they should continue.” This time, they worked with an order, but when they were not volunteering, it was questionable how good the exams were. Then I changed them without telling the rector, this time we assigned the new recruits but could not send them to training. Some problems are beyond us, YÖK is not aware of them. Someone has to explain these issues somewhere in a suitable language.

One testing member in another university explained further:

Ben 29 yıllık hocayım ama şu son üç senede testingin bana çok şey kattığını düşünüyorum. Ve keşke testingde daha önce çalışsaymışım diye de düşünüyorum. Her sene, bir önceki sene yaptıklarımın daha iyisini yaptığımı hissettim kendi adıma. Ama bizim durumumuzda, kimse 2-3 yıldan fazla çalışmak istemez. Bu benim üçüncü yılım artık daha

fazla yapamayacağımı düşünüyorum. Ama şartlar iyileştirilse “sınav yetiştirebilecek miyim” kaygısı yaşamadan devam edebilsem, devam etmek isterdim testingde. Az derse girerek çoğu zamanımı sınav hazırlamaya ayırmak ve bu konuda kendimi geliştirmek isterdim. Buraya gelen zaten ben 2-3 sene sonra gideceğim, başkası gelecek diye düşündüğünden sistem geliştirmek, kalıcı bir şey yapmak için çaba göstermiyor. Anlatabiliyor muyum? Belki amiyane ama zaten çıkacağım testingden, kalıcı değilim dediğim zaman da eğitim almak çok gerekli gelmiyor bana. Belli bir noktadan sonra daha profesyonel bakmaya başlıyorsunuz. Bir de arkadaşımın örnek vereyim; A... ilk başladığında testinge “burada istediğimiz kadar çalışabiliyor muyuz?” diye sormuştu, bu yıl kaçmaya çalışıyor. Kaçamadı, bir yıl daha burada.

I have been an instructor for 29 years, but I think that testing has contributed a lot to me in the last three years. And I wish I had worked in testing earlier. Every year, I felt that I was doing better than I did the previous year. But in our case, no one wants to work more than 2-3 years. This is my third year now I think I can't do anymore. But if the conditions were improved, if I could continue without worrying about “will I be able to complete the exam?”. I would like to continue testing. I would like to spend most of my time preparing for exams by attending few classes and improving myself in this regard. The members beginning to work here may think that they will go after 2-3 years and someone else will come, so they do not make an effort to develop a system and do something permanent. Am I clear? When I feel I'm not going to be permanent, I don't think it's necessary to get training. After a certain point, you start to act more professionally. Let me give an example from my friend; When A... started to work in the testing unit, she asked “can we work here as much as we want?”, but this year she is trying to escape. She couldn't yet, and she's here for one more year.

Many TUMs indicated in the interviews that they were concerned about the need for support staff (N=230). Although there are technological devices and online tasks are implemented in testing units, the absence of an expert in charge of technical affairs in these units may cause challenges from time to time. What is worse, when a problem is encountered on a device, their efforts to solve the problem with their own efforts, since it takes a long time to wait for an expert from the company or institution itself, which leads to a waste of time, actually cause bigger problems. A testing member, for instance, put an emphasis on irrelevant work and stated:

Bir listening sınavında teknik bir sorun, teknik bir arıza çıktığında biz ilgileniyoruz. Biz teknik personel değiliz ki. Niye CD, USB, bilgisayarla ben uğraşayım. Ben de bilmeyebilirim, ben bunu bilmekle yükümlü değilim ki. Ben de teknik açıdan illiterate olabilirim. Bana teknik konularda böyle bir eğitim verilmedikten sonra kendi kişisel çabamla çözmeye çalışıyorum. Mesela şurada duran 8000 kişinin girdiği Erasmus sınavı

kağıtlarını şu optik makinede biz dördümüz bütün gün okutuyoruz mesela. Sonuçları alıyoruz, makinede değerlendiriyoruz. Sürekli yanlış kullanımdan dolayı arıza çıkıyor, çözemiyoruz da. Fotokopi odasında bir memur var. Geçen bilgisayarla yazıcının bağlantısı ile ilgili bilgisayarımda bir sorun çıktı mesela, kendisini çağırdım ama çözemedi. Sonra eşimden rica ettim o uzaktan bağlandı da yaptı. Yani böyle kişisel çözümler üretmek zorunda kalıyoruz. Adamın da sorumluluk ya da görev tanımı olmayınca bir şey diyemiyoruz.

When a technical problem arises in a listening exam, we should deal with it. However, we are not technical personnel. Why should I deal with CDs, USBs, and computers? I may not know, but I am not obliged to know, either. I might be technically illiterate. Although I have not been trained in technical issues, I am trying to solve them with my own personal effort. For example, all day, we, four of us, scan the Erasmus exam papers of 8000 students in this optical machine. We note the results and evaluate them on the machine. Due to constant misuse, a malfunction occurs, and we cannot solve it. There is an officer in the copy room. For example, there was a problem on my computer related to the connection of the printer with the computer. I called him, but he could not solve it. Then I asked my wife and she connected remotely and solved it. Therefore, we have to produce such personal solutions. We can't say anything to him because he doesn't have such a responsibility or job description.

Another director drew the attention to another side of the issue:

Bu kağıtların hepsinin şeyini düşünün, optiği, zarfı, güvenliği, taşınması. Mesela dosyalar var, kocaman dosyalar, taşınacak bir yerden bir yere. Görevli yok, biz taşıyoruz. Başka yerden hizmetli çağırıyoruz, bir hafta bekliyoruz, gelene kadar kendimiz yapıyoruz. Şu karşı oda komple arşiv, sınavlar üst üste. Onları kendimiz paketledik tek tek, o koca kolileri. Arkadaşımız hastaneye gitmek zorunda kaldı. Bize destek elemanları gerekli. Evet, fiziksel desteğe ihtiyacımız oluyor. Hizmetli arkadaş her zaman gelmiyor, 7/24 başka yerde. Onu bekleyene kadar hadi kendim yapayım diyorsun mecburen.

Think about all the work of these papers; the optics, envelope, security, and carrying. For example, there are files, in huge packs, to be moved from one place to another. There is no officer, we should carry them. We call staff from another place, we wait for a week, and we do it ourselves until one comes. That opposite room is the archive room, the exams were in a mess. We packed those big boxes ourselves, one by one. Our friend had to go to the hospital for a week. We need support staff. Well, we need technical support for physical work. A technical staff does not always come, but elsewhere 24/7. Instead of waiting for staff, we solve it ourselves.

A number of participants indicated in the interviews that there is a lack of orientation (N=202) once they start to work in the office. As mentioned earlier, there is frequent staff circulation in TUs. What is more, all TUMs, especially in state universities, change at the same time, which may push the newly appointed members into an ocean without any fins if there is no orientation or training programme. A testing member cited her concern about the importance of orientation:

Ofise başladığımızda zaten hizmet öncesi eğitim alamıyoruz, bari bir oryantasyon eğitimi olsa. Maalesef, önceden burada çalışanlar buradan kurtuldukları için buranın önünden bile geçmek istemiyorlar. Sıfırdan işlerin öğrenilmesi zor oluyor. Hele bir de geçen sene tüm ekip aynı anda değişmişti. Geçen yılki sınavlar çok eleştirildi. Sistemli bir oryantasyon programı olsa çok iyi olurdu.

When we start to work in the office, we can't get pre-service training anyway. I wish there were an orientation training. Unfortunately, those who used to work here don't even want to pass in front of this place after they got rid of this office. Learning things from the beginning is difficult. For instance, last year, the whole team changed at the same time. Last year's exams were heavily criticised. It would be great if there was a systematic orientation programme.

Among the challenges, ambiguity in the operational system (N=176) was a great concern for testing members. Although quality studies have progressed in THE system, it is difficult to claim that it has a significant effect on the processes of preparatory schools, especially on examination, when their reflections on the university are considered. It was especially emphasized that there are no permanent solutions and a system has not been established in many of the institutions interviewed which causes unclear workflow. One testing member expressed her feelings about the ambiguity in these words:

Sürekli sistem değiştiği için biz de aslında iş akışını oturtmaya çalışıyoruz. Biz de çok bilemiyoruz ama henüz bir sistem oturtma aşamasında olduğumuz için daha da fazla yoruluyoruz bu belirsiz tempodan. Bunun bir sebebi de hem sistem değişikliği hem de akreditasyon süreci aynı ana denk geldiğinden.

Since the system is constantly changing, we are actually trying to establish our workflow. We don't know much either, but since we are still at the stage of setting up a system, we get even more tired of this uncertain pace. One reason for this is that both the system change and the accreditation process have coincided at the same time.

A testing member added some more details and elaborated on the lack of clarity in their workflow as follows:

İş akış şeması net olmadığından kim denk gelirse o anda, o mevcut işi yapıyor. Sınav hazırlama sürecini gösteren genel bir takvimimiz var, baskı, şu, bu. Seviye paylaşımını da aramızda soruyoruz “A1 mi istiyorsun, sen A2 mi istiyorsun.” Bazen de değiştiriyoruz, mesela ben bu sınavda writing yapayım, sen bir sonrakini al. Kendi aramızda böyle yapıyoruz, öyle strict bir şey yok. Mesela ben şu an optik yaptım, S... zarf yazdı. Bir dahakine ben zarf yazarım, o optik yapabilir. Yani bu benim işim değil, beni ilgilendirmez demiyoruz yani. Kim müsaitse. Bugün baskıyı kim yaptı, ertesi gün paketlemeyi başka biri yapıyor. Bu sorun olabiliyor bazen. İşler karışabiliyor, son dakika çözüm bulmamız gerekebiliyor ve günü kurtarmak zorunda kalıyoruz.

Since the workflow chart is not clear, whoever comes across is doing the existing job. We have a general calendar showing the exam preparation process, edition, and so on. We also ask each other about level sharing, “Do you want A1 or do you want A2?”. Sometimes we change it, for example, I can prepare the writing part in this exam, and you take the next one. This is how we do it among ourselves, there is nothing so strict. For instance, I'm currently doing the optics and S... is writing the info on the envelopes. Next time I will be writing the info on the envelopes, he can do the optics. Thus, we do not object saying it's not my business and it does not interest me. Whoever is available does. When one does the printing today, someone else does the packaging the next day. This can be a problem sometimes. Things can get messy, we may have to find a last-minute solution and we have to save the day.

In addition to the concerns stated above, the need for a comfortable work environment (N=124) is a factor contributing to anxiety. As mentioned earlier, there are poor physical conditions in most of the universities where only one or not even one office is left for TU. Considering all devices, including copy machine, optic, and exam documents, including previous papers along with resources and books are placed in the unit, TUMs claim to feel demotivated, as shown in the interviews. Besides, once a member has a meeting with an instructor or level coordinator, another might be trying to write the distractors. As explained by especially testing members the participants:

Çalıştığımız oda çok dolu ve karmaşık. İçerisinde kağıtlardan tutun da arşiv kağıtları, materyaller, bilgisayarlar, optik formu, fotokopi makinesi gibi bir sürü şey var. Gördüğümüz gibi dolaplarımız çok yeterli olmadığından dokümanları yerlere de koymak zorunda kalıyoruz. Tüm olay burada gerçekleşiyor, birisi fotokopi çekerken öbürü orada arşivle ilgileniyor, öbürü odada sınav sorusu yazmaya çalışıyor mesela. Benim proofreaderım geldiğinde N... soru yazmaya çalışıyor ama ben proofreaderımla konuşmak

zorundayım mesela. Öyle bir görüşme odamız da yok. Mesela baskı yapıldığında ben nasıl konsantre olabilirim. Burada fotokopi makinesinin sesinden oturup soru yazmak işkence oluyor. İster istemez bizi rahatsız ediyor. Bir sınav bastığımızda üç gün boyunca bu makinenin sesi bu minnak odanın içinde mesela.

The room we work in is full and complex. It contains a lot of stuff, from papers to archival papers, materials, computers, optical forms, and photocopy machines. As you can see, our shelves are not very sufficient, so we have to put the documents on the floor. The whole process takes place here, while someone is making a copy, another is there dealing with the archive, and the other is trying to write an exam question in the room. For instance, when my proofreader comes, N... is trying to write a question but I have to talk to my proofreader. We do not have a meeting room. How can I concentrate when the printing is being conducted? It's like torture to sit down and write questions over the sound of the copy machine here. It inevitably bothers us. For example, when we print an exam, the sound of this machine lasts three days in this tiny room.

Another testing member added:

Yaşadığımız en fazla sorunlardan biri de buradaki trafiğin hiç bitmemesi. Hoca geliyor kağıtları bırakıyor, diğer hoca zarf almaya geliyor, öğrenciler geliyor sınavda ne çıkacak diye soruyor, itiraz dilekçesi bırakmak istiyor. Ofiste aslında bitmeyen bir trafik var. Sınav güvenliğinden gürültüye kadar.

One of the biggest problems we have is that traffic in the office never ends. The teacher comes and leaves the papers, another teacher comes to get envelopes, the students come and ask what will happen in the exam, and leave their objection forms. There is actually never-ending traffic in the office. From exam security to noise.

Some of the testing members approached the challenges by emphasizing ineffective coordination among testing members (N=89). Due to the lack of a systematic workflow and job description, it was stated in the interviews that some ego clashes would emerge among the members when there is no harmony among TUMs. As a result, they prefer to work individually rather than in teamwork. One of the testing members specified her concern on this issue as follows:

Sınavın hazırlanmasından uygulanması hatta sonrasındaki süreçlerde koordinasyon çok önemli. Biz aslında hani doğruyu söylemek gerekirse sürekli birbirimizle yardımlaşmıyoruz, akıl danışmıyoruz, sürekli fikir alışverişi yapamıyoruz. Bizden önceki grupta çok iyi bir uyumumuz vardı, ama arkadaşlarımız ayrıldı. Şimdi ise çok bireysel çalışıyoruz. Bu koordinasyon eksikliği motivasyonumuzu da aslında kırıyor.

Coordination is very important in the preparation, implementation, and even following processes of the exam. In fact, to tell the truth, we can't always help each other, consult each other, and exchange ideas. We used to get along well in our previous group, but our friends left. Now we work rather individually. This lack of coordination actually affects our motivation.

A testing member shared the same perception and added some more details:

Ekip içerisinde uyum yoksa kişisel olaylar başlıyor. A... bana bunu yaptı, o da bana böyle yapmıştı falan. X... Üniversitesi'nde böyle bir sıkıntı yaşadık. Bir hocamız, sadece şu bölümü hazırlarım diyordu, ona başka bir bölüm veremiyorduk. Ayrıca, koordinasyon ve uyum sorunu olduğunda birim içerisinde sürekli ego çatışması oluyordu. Mesela, bir kadın hocamız kendi yaptığı sınavın değiştirilmesinden rahatsız oluyordu. Belli bir düzen olmayınca testing member lar arasında çok büyük bir fark var.

If there is no harmony in the team, personal issues begin. A... did this to me, that member did that to me, etc. We had such a problem at X... University. One of our members was saying that she would prepare only a specific section, we could not give her another section. In addition, there is constant ego conflict within the unit due to a lack of coordination and harmony. For example, a member was uncomfortable with the revision of her part. In the absence of a certain order, there is a huge difference among testing members.

Finally, testing members rarely mentioned the problems with the previous system (N=66). Most of the TUMs complained about the continuity of the challenges despite their personal efforts, which is believed to take place due to enduring problems passing from one year to another. One of the testing members explained in the following extract:

Yönetim değişiyor, birimdeki arkadaşlar değişiyor, ancak değişmeyen bir şey var o da yaşadığımız sıkıntılar. Ulusal otoritelerin yapmadığı düzenlemeler bizi etkiliyor, ancak alt düzeyde mevcut şartlarda bir sistem geliştirebiliriz ki bizden sonra gelen arkadaşlar en azından ne yapacaklarını bilsinler. Ancak önceki testing member lar sadece günü kurtarmışlar. Düzensiz, kayıtsız, belgesiz biçimde sınavlar hazırlanmış, uygulanmış ve arşivlenmiş. Geçmişe dönemiyoruz, dosyalama yok. Yeni gelen ekip okyanusta balık arıyor adeta.

Directors have been changing, so have testing unit members, but there is one thing which does not change and that is the problems we experience. The lack of regulations by national authorities affects us, but we can develop a system under current conditions at the unit level so that other members who work after us can at least know what to do. However, previous testing members only saved the day. Irregular, unregistered, undocumented exams were

prepared, administered, and archived. We can't go back, no filing. The newly arrived team is like looking for fish in the ocean.

The results reveal that there are problems on the unit basis, depending on the problems experienced on a national and institutional basis or due to other reasons. The most striking of these problems is heavy workload and insufficient number of TUMs. These problems are followed by excessive TUM turnover and a lack of support staff. The next subsection deals with stakeholder-related challenges.

4.1.2. Stakeholder-related challenges

Stakeholder-related challenges comprise the second most frequently stated type of challenges (N=4.804; 27,85%) encountered in testing practices at EPPs in higher education. This category consists of the challenges related with administrators, testing members, instructors, and students.

Table 4.6. *Stakeholder-related challenges*

Sub-categories	Number of Codes	%
Challenges related with administrators	1.221	25,42
Challenges related with testing unit members	1.706	35,51
Challenges related with instructors	778	16,19
Challenges related with students	1.099	22,88
Total	4.804	100

As Table 4.6. presents, the highest-ranking subcategory belongs to challenges related with administrators consisting (N=1.221; 25,42%). The second highest-ranking subcategory is challenges related with testing members (N=1.706; 35,51%). Challenges related with instructors (N=778; 16,19%) and with students (N=1.099; 22,88%) are given in the last two categories. The Figure 4.3. shows the sub-categories and their codes of the stakeholder-related challenges.

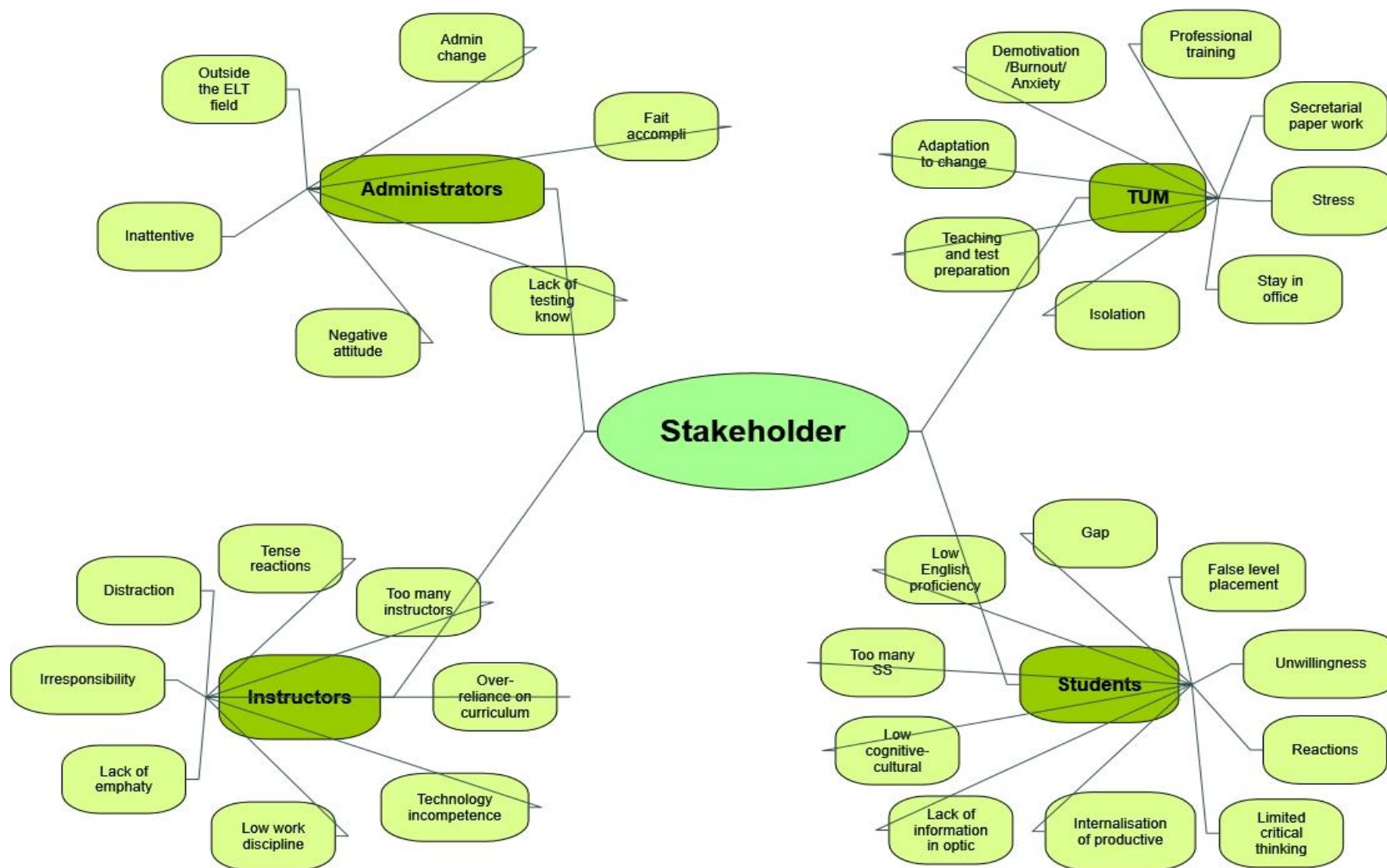


Figure 4.3. Stakeholder-related challenges

The next subsections deal with the challenges related with administrators, with testing unit members, with instructors, and with students, explained respectively.

4.1.2.1. Challenges related with administrators

This subcategory includes six emerging codes which are shown in Table 4.7.

Table 4.7. Challenges related with administrators

Emerging Codes	Number of Codes	%
Fait accompli	296	24,24
Administrators change	264	21,62
Administrators from outside ELT	216	17,69
Inattentive administrators	178	14,58
Lack of testing knowledge	172	14,09
Negative attitudes	95	7,78
Total	1.221	100

The most frequent anxiety in this subcategory was testing members' concerns about fait accompli (N=296). It is stated that hard work and excessive workload prevent these offices from being attractive. In this case, directors tend to make forced assignments. Therefore, the exams can be of poor quality and TUMs work without motivation and feel lonely. After a while, they attempt to leave the office as it is time that a new assignment is made by force again, which goes on a cycle. A director mentioned in the interview:

Tabii ki ölçme değerlendirme ofisine çok fazla başvuru bugüne kadar gelmedi. İş yükünün biraz fazla, işlerin zor olduğu için, biraz da çok cazip bir ofis olmadığından son üç yıldır bu ofise başvuran olmadı. Diğer ofislerde çalışmak istemiyorlar ama seçim yapacaklarsa diğerlerini seçiyorlar. Biz de ya zorla ya da ikna yöntemiyle testinge görevlendirme yapıyoruz. Öyle bir yetkimiz aslında yok tabii ki, ayrıca zoraki görevlendirme de etik değil ama yapmak zorunda kalıyoruz. Görevlendirdiğimiz arkadaş kalitesi düşük sınav çıkarttığında yaptırımımız yok. Ofisten göndermek de çözüm olmuyor. O kişiyi görevden aldığınızda o pozisyonda tekrar ihtiyaç ortaya çıkacak. Bazı özellerde çalışanlar iş güvencesini bir tehdit olarak algılayıp bu ofislerde elinden geldiğince iyi çalışıyorlar ama devlette olmuyor.

Of course, not many instructors have applied to the testing unit yet. Because the workload is a bit too much, the work is difficult, and it is not a very attractive office, no one has applied to this office during the last three years. They don't want to work in other offices,

but if they're going to choose one, they choose other offices. We assign testing members by force or persuasion. Of course, we do not have such authority, and fait accompli is also unethical, but we have to do it. When the staff we appoint prepares a less quality exam, we have no sanction. Sending him from the office is not a solution, either. When you dismiss that person, a need for that position will arise again. The members working at some private universities perceive this assignment as a threat to job security and work in these offices as best as they can, but this is not the condition in the state ones.

A testing member explained the issue in a more detailed way:

Geçe sene müdürümüz, arkadaşlar böyle iki tane birim oluşturacağız: curriculum ve testing. Gönüllü olmak isteyen var mı diye sordu. Curriculum birimi hemen doldu, ancak testing birimine önce birkaç kişi gönüllü oldu, bir hafta sonra karar değiştirdiler. Yapmak isteyip cesaret edemeyen arkadaşlar vardı. Müdür bey de yapabileceğini düşündüğü bu arkadaşlara “sen yapabilirsin” diye ikna etmeye çalıştı, ancak genel kitle “hocam ben yapamam” dedi, sadece bir kişiyi ikna edebildi. Sonra zoraki bir görevlendirme yapıldı. Zorla görevlendirilenlerden hamile olan arkadaş “benim için gerçekten çok stresli, yani uyuyamam” diye itiraz etti. Hocam dürüst olayım zorla görevlendirilenlerden biri de benim. Yani ben testinge girdim ama çıkamadım hocam. Mahsur kaldım. Geliş şeklim de garipti, ders bitince buraya geldim gibi bir şey oldu.

Last year, our director mentioned we were going to create two units: curriculum and testing. He asked if anyone would like to volunteer. The curriculum unit was quickly filled, but initially few volunteers volunteered for the testing unit, but they changed their minds a week later. There were friends who wanted to do it but could not dare. Our director tried to persuade those friends, who he thought they could do, encouraging “you can do it,” but most of them rejected “I can't do it,” and he was able to convince only one person. Then a forced assignment was made. “It's really stressful for me, so I can't sleep.” objected the pregnant member through a forced assignment. To be honest, I am one of those who were forced to work. That is, I started to work on testing, but I could not get out. I have been stuck. The way I began was also strange, it was like I came here just after my class was over.

A testing member shared similar views but also put emphasis on the instant assignment:

Aslında bizim görevlendirmemiz şu şekilde oluyor, ilk önce yönetim birimlerde çalışmak isteyen var mı diye bir mail atıyor. Testing birimine hiç kimse cevap vermiyor yani hiç kimse çalışmak istemiyor. Mesela geçen sene kimse çalışmak istemeyince biz bireysel olarak çağrılıp, çalışır mısınız diye sorulduk. Yeni başlayanlar olarak birazcık zoraki kabul ettik bu görevi. Bizim tam olarak ne zaman başladığımız bile belli değil. Ben nisanda benim

elimden bazı derslerim alınarak buraya geldim mesela. Buraya bir anda geldiğimden diğer derslerim bölüştürüldü diğer hocalar arasında. Testing ofiste çok ihtiyaç var diye buraya geldim. Testingde çalışırken bir yandan kalan dersin şeylerini de devam etmek zorunda kaldım: çocukların coursebook projeleri, writingleri... Zoraki olarak hem burada çalıştım hem onlara baktım bir yandan da.

In fact, our assignment is as follows, first of all, our director sends an e-mail to see if there is anyone who wants to work in the units. No one volunteers for the testing unit, that is no one wants to work. For example, last year, when no one wanted to work, we were called individually and asked if we would work. As new staff, we were required to accept the offer. We don't even know when exactly we started. For example, When I started to work here in April, some of my lessons were taken from me. Since I came here all of a sudden, my lessons were assigned to other instructors. I came here because there was a great need in the testing office. While I was working in the testing office, I had to continue to do the work for my remaining courses: such as students' coursebook projects, and writing papers... I worked here and did them at the same time by force.

The challenges related to administrator change (N=264) comprise the second most frequently used type in this subcategory in the interviews. The constant change of directors prevents the internalisation of the problems encountered in Tus. Newly appointed directors, on the other hand, may realise these problems too late as the problems are growing exponentially. A testing member, for instance, reported:

Üniversitede Rektör değişiyor, bizim müdür de değişiyor. Müdürümüz her değiştiğinde yeni gelen müdüre yaptıklarımızı anlatmak, hatta ikna etmekle uğraşıyoruz. Siz neden az derse giriyorsunuz diyor, aslında ben sadece dört saat az derse giriyorum. Bırakmak istediğimizi söylediğimizde ise müdür yardımcıları izin vermiyorlar. Yeni gelen müdür süreçlerimizi, işlerimizi öğrendiğinde görev süresi bitmek üzere oluyor. Her yönetim değiştiğinde bizim sistem de etkileniyor, bizim birim de sistemsel değişiklikler yaşıyor. Zaten düzenli bir sistem yok, olan da heba oluyor.

When the rector changes at our university, our director also changes. Every time our director changes, we try to explain and even persuade the new director about what we are doing. The new director asks why we attend fewer classes. In fact, I have only four hour-reduction. When we say we want to quit, our vice director does not allow us. When the new director gets familiar with our procedures and work, the term of her/his office is about to end. Every time the director changes, our system is also affected, and there are systemic changes in our unit. There is no regular system anyway, then everything is wasted.

The problems resulting from the directors outside the ELT field (N=216) have the third highest frequency among the challenges in this subcategory testing members face when codes are computed. It was stated in the interviews that proximity to the rector, professor title, and rewarding appointments sought in director appointments rather than knowledge of preparatory education prevent the prioritisation of the problems. A testing member elaborated on the issue as follows:

Bizim alandaki hocalarımızda yüksek akademik title olan çok yok, bu yüzden müdürlerimiz genelde alan dışından oluyor. Bu kişiler de başka alanda mesela profesör oluyor. Hoca profesör olabilir ama alan dışı olunca çok fazla sorun yaşıyoruz. Okulumuza alandan doğru kişiler atansa daha doğru kararlar alınabilir. YÖK'ün yaptığı projelerde alan uzmanlığının önemi vurgulanıyor ama bizim görevlendirmelerde pek alan uzmanlığı aranmıyor maalesef.

Most of the academics in our field do not have high academic titles, so our directors are usually from outside the field. They are professors in another field, for example. A person can be a professor, but we have a lot of problems when our director is out of the field. Better decisions can be made if the right people from the relevant field are appointed to our school. The importance of field expertise is emphasized in the projects carried out by YÖK, but unfortunately, field expertise is not sought in our assignments.

Another member explained further:

Son üç müdürümüz de mühendisti, ondan önceki de su ürünlerinden. Bizden önce bir hocamız ilk defa ELT'den bir müdürümüz olmuş. Geliştirilen iyi uygulamaların çoğu o döneme dayanıyor. Sonrasında standart edinmeyi, kalite çalışmaları için çok uğraştık ancak gerekli desteği bulamadık. Çok büyük bir sorun aslında. Şimdi ELT tabanlı biri olsa, daha fazla işin içine dahil olurdu. Ama çoğu üniversitede, özellikle devlet üniversitelerinde, öyle değil hocam maalesef.

Our last three directors were engineers, and before them was one from fisheries. A director before us was from ELT for the first time. Most of the good practices date back to that period. We tried hard to obtain standard and quality studies, but we could not find the necessary support. It's a huge problem actually. If our director were ELT-based, he'd be more involved. But in most universities, especially in state universities, this is not the case, unfortunately.

Another major concern of the directors and testing members related to national challenges is inattentive directors (N=178). Most TUMs stated that directors might be indifferent to their problems in the office owing to the fact that the problems are

complex and their source is related to higher-level authority institutions. One of the testing members explained her concerns about their inattentive manners as follows:

Yöneticiler, bizim sorunlarımızla çok uğraşmak istemiyor, değişiklikle ilgilenmek istemiyor. Geçen bir şey talep edildiğinde “zaten arkadaşlar yapıyor, değişikliğe gerek yok” demiş. Yapıyoruz da gece mi çalışmışız, evde mi çalışmışız, çok bir anlamı olmuyor. Mesela, bu kitapların hepsi çok eski, senelerdir rafta olan kitaplar ve tükenmiş durumda artık içleri. Bunu idareye 2-3 senedir, yani biz buraya geldiğimizden beri aktarıyoruz. Yeni kaynaklar alınmasını ricada bulunuyoruz, ama bir sonuç yok.

Our directors do not want to deal with our problems, they do not want to make a change. When something was requested, our current director said “friends are already doing it, there is no need for change.” We manage it, but whether we worked at night or at home is not taken into consideration. For example, all of these books are very old, on the shelves for years, and now out of use. We have been conveying this issue to the administration for 2-3 years, that is since we came here. We are asking for new resources, but no change.

Another testing member added:

İhtiyaçlarımızı, sorunlarımızı müdüre iletiyoruz ama ondan sonra herhangi bir aksiyon alınmıyor. Geçen hafta sınav güvenliği ve riskleri ile ilgili durumu aktardık. Sınav masanın üstünde kalıyor, “yangın çıksa” diyorum, “yansa iptal edeceğiz, canımdan kıymetli mi” diyor.

We explain our needs and problems to the director, but after that, no action is taken. Last week, we informed him of the exam security and risks. I told, “the exam papers stay on the table” and warned him if there was a fire. He said, “we can cancel it if it burns, is it more valuable than my life?”.

Some of the testing members indicated that the lack of testing knowledge of the directors leads to the continuity of the challenges (N=172). It is emphasized that the director must believe that the challenge in a TU is important to solve considering that the problems in the administration seem to be more concrete for them. Otherwise, the director who has insufficient knowledge of this subject does not want to deal with it. One testing member focused on the issues related to the lack of testing knowledge of her director:

Ben ilk testinge başladığımda soruyorum, soruyorum, ancak hocam yöneticiler sorulara cevap veremiyor. Item analysis nedir diyorum, önemliyse yapın diyor. Tabii hoca nerden bilsin eğer testing biriminde çalışmadıysa. Hoca profesör, ancak alandan olması ve profesör olması yetmiyor. Sen bir şey söylüyorsun, hoca tamam yapalım diyor ama hoca ne olduğunu bilmiyor. Hani öyle şeyler söylüyor ki testingin “t” sinden anlamadığı belli.

When I first started testing, I asked and asked, but our administrators could not answer my questions. When I asked what item analysis was, he said if it was important, we could do it. How would the director know if s/he did not work in the testing unit? The person can be a professor. It is not enough that the director is from the field and a professor. We are explaining something, and the director agrees to do it, but s/he does not know what it is. S/he says such things that it is obvious s/he does not understand the “t” of testing.

One final reason why the participants get worried in the unit seems to be the negative attitudes of the directors (N=95). It was mentioned that the directors coming from other fields and having limited or no knowledge about testing could make them more result-oriented. In this regard, they can be easily affected by complaints and exhibit a negative attitude towards TUMs. One testing member explained how she felt:

Sınav biriminde çalıştığımızdan başarısızlığın, dikkatin hepsi üstümüzde. Öğrenciler sınavdan kaldığında bizi idareye ve hocalarına şikayet ediyor. Hocalar sınavdan kaldığında bizi hedef biliyorlar. Daha kötüsü ise yönetim bizi uyarıyor, bize negatif davranıyor. Hem ek kazancım yok hem de sürekli yönetimin hedef tahtasında oluyorum. Mesela geçenlerde Türk Hava Yolları'nın eğitimi vardı, bu pilotlara sertifika eğitimi verilecekti, ona gidecektik. Müdür bey ilk önce önceliği bize verdi, D... ile ikimiz gidecektik. Bu durum hocaları çok rahatsız etmiş, sürekli bizi şikayet etmişler. Hocalar kendilerinin bir sürü derse girdiklerini, bizim ise bir tek ofisten çalıştığımızı söyleyip müdürün kafasına girmişler. Normalde müdür beyin bakış açısı biraz daha farklı, ofislerle ben bu sistemi yürüteceğim diyordu. Ama tepkiler ve şikayetler sonrasında tavırları değişti, ve eğitime biz gidemedik. Hocalara siz de o zaman ofise gelin çalışın diyoruz, ama insanlar pek gönüllü olmuyorlar.

Since we work in the testing unit, all the failure and attention are focused on us. When students fail the exam, they complain about us to the administration and their instructors. When instructors' classes fail, they determine us as a target. Worse, the directors warn us and treat us negatively. I do not have any additional income and I am constantly on the target of the management. For example, there was a training in Turkish Airlines recently, a certified training was going to be given to pilots, and we would lecture in the training. Initially, the director gave us priority, so I was going to participate with D.... This situation disturbed the teachers very much, they constantly complained about us. Then they influenced the director by saying that they are attending many classes but we only work in the office. Normally, the director's point of view was different as he said that he would run the system with the offices. But following the reactions and complaints, his attitude changed, and we could not go to the training. We tell the instructors that they can come to the office and work then, but they do not volunteer.

Challenges related with administrators seemed to have an impact on other challenges. The most frequently mentioned challenge was the way of assignment in the testing unit. It was claimed that although they were reluctant to work in TUs, they were assigned through *fait accompli*. Besides, administrator change, the field, inattentive and negative attitude, and limited knowledge on testing of the directors were other major challenges influencing TUMs' working conditions and motivation in the office. The next subsection deals with challenges related with testing unit members.

4.1.2.2. Challenges related with testing unit members

This subcategory includes eight emerging codes which are shown in Table 4.8.

Table 4.8. Challenges related with testing unit members

Emerging Codes	Number of Codes	%
Demotivation	404	23,68
Need for professional training	363	21,28
Secretarial paperwork	311	18,23
Stress	250	14,65
Need to stay in office	135	7,91
Need for isolation	97	5,69
Teaching and test preparation at the same time	80	4,69
Adaptation to change	66	3,87
Total	1.706	100

Demotivation (N=404) has the highest frequency among the challenges in this subcategory testing unit members face when codes are computed. Excessive workload, time constraints, and high tempo due to the high number of students and exams in a term force TUMs to work under pressure without any additional gain. Moreover, instructors' as well as directors' ungrounded comments and negative attitudes make them feel rather anxious, leading to demotivated and burned-out in the end. A testing member explored as follows:

Sınav temposu çok yüksek oranda olduğundan motivasyonumuz etkileniyor. O kadar önemli ki motivasyon. Maddi manevi motivasyon kaynağımız yok. Tamam iç motivasyonumuz var, intrinsic motivasyonumuzla bir şekilde yapıyoruz ama insan bazen sürekli risk altında, sürekli alert, tetikte beklemeye gerek var mı diye düşünüyor. Bir sonraki sınavda sorun olacak mı, olmayacak mı, güvenlik önlemi açısından bir problem olacak mı diye

endişeleniyoruz. Metrobüs kazası oluyor bizi etkiliyor, başka bir olay oluyor yine bizi etkiliyor. Bu kadar şeyi düşün düşün, sürekli ek çalışma, ekstra işler yap ama bir karşılığı yok. Bunun yanı sıra sınavlara gelen yorumlar da moralimizi bozuyor. Nasıl diyeyim bu işten gerçekten anlayan biri solid bir yorum yapsa evet o zaman derdim 'evet ben yapamamışım' ama gelen yorumlar açıkçası çok kişisel, sübjektif yorumlar olduğu için biz çok etkileniyoruz açıkçası. O kadar emek harcayıp bir de haksız eleştirilince tabii ki moralimiz anlık bozabiliyor. Çok sübjektif eleştiriler yani tamamen kendi algılarına göre. Öneri yok hiçbir zaman, sadece olumsuz bir eleştiri var.

Since the pace of the exam is very high, our motivation is affected. Motivation is so important. We do not have a financial motivation source. Well, we have internal motivation, we do it somehow with our intrinsic motivation, but sometimes we think that we are always at risk, constantly alert, and ask whether there is a need for that. We are worried about whether there will be a problem in the next exam or if there will be a problem in terms of security measures. A Metrobus accident happened that affected us, another event happened that affected us again. We always think about all these things, constantly doing additional work, extra work, but there is no reward in return. In addition to this, the comments on the exams also upset us. If someone who really understands these issues makes a solid comment, then yes, I will say "yes, I couldn't do it," but frankly, we are very impressed because the comments are very personal and subjective. Of course, when we spend so much effort and are unfairly criticised, we can lose our motivation instantly. Subjective criticism, that is, completely according to their own perceptions. No suggestions at all, just a negative review.

Another testing member added:

Ofiste çalışanlarda ciddi burnout etkisi oluyor, o etkiyi hafifletmek için son iki yıldır bir rotasyona gitmek zorunda kaldık. Burada çalışacak kişiler kurban gibi bekliyor. Çünkü çalışma temposu çok yüksek, sınav dönemleri gece çıkıyorsun sabah erken yine geliyorsun aslında. Benim iznim birikti mesela izne çıkamıyorum. Hocam şu an 40 gün iznim var, sizinle görüşmeye buraya gelmek için zar zor izin aldım. 40 gün iznim olmasına rağmen bir de yalvarıyorum izin almak için yani hakkım olan şey için niye yalvardığımı anlamıyorum hocam. İşte testere filminin sekizinci bölümünü burada çekmek istiyoruz. O kadar ümitsiz.

There is a serious burnout effect on the ones working in the office, we have had to go on a rotation for the last two years to alleviate that effect. The people who will work here are waiting like victims. Because the work pace is very high, we leave here at night during exam periods, you come back early in the morning. My leave has accumulated, for example, I cannot go on leave. Sir, I have 40 days of leave right now, I barely got permission to come here to participate in the interview with you. Although I have 40 days of leave, I beg for permission, I do not understand why I am begging for what is my right. This is where we want to shoot the Saw 8 since we are so hopeless.

A director shared similar views but also put emphasis on their anxiety level:

Öğrenci sayısının fazla olması, testlerdeki beklentinin çok yüksek olması, en ufak hatada suçlanabilmeleri biraz onları da zorlayan, endişeye sevk eden durumlar. İşleri yetiştirmek için bu sefer ek mesai yapmak zorunda kalıyorlar, bu sınav bitiyor diğerinin hazırlığı başlıyor. Sürekli endişe ve kaygı ile ofise geliyorlar. Hem insanların, hocalar, öğrenciler ve bazı yönetimler, dozu kaçan biçimde hedefi olabiliyorlar, bir yandan da zamanla yarışıyorlar. Sonucunda da onlarda ciddi kaygılar oluşuyor.

The high number of students, high expectations in the tests, and the fact that they can be blamed for the slightest mistake create tough times for them and cause anxiety. Then they need to work overtime to complete the work. Once one exam ends, the preparation for the other begins. They come to the office with constant worry and anxiety. Instructors, students, and some directors can put them on target. In addition, they are competing with time. As a result, they have serious concerns.

The second most frequent anxiety in this subcategory was testing members' concerns about the need for professional training (N=363). Exam preparation requires a lot of technical knowledge as well as general grammatical and pedagogical knowledge. However, it was shown that most of the TUMs in these units did not have any training other than the 3-credit testing course in undergraduate education or half-day seminars. One testing member stated in the interview:

Geniş bir kitleye sınırlı sayıda personelle sınavlar hazırlıyoruz, ancak bu konuda ciddi eğitimlere ihtiyacımız var. Mesela sınav hazırlama aşamasında item writing konusunda bir eğitim almadık, ayrıca sınavın sonucunda standardizasyonun nasıl sağlanması konusunda da yeterli bilgimiz yok. Şimdiye kadar kendimizi alaylı hissediyoruz.

We are preparing exams for a number of students with a limited number of members, but we need serious training on testing. For example, we did not receive any training on item writing during the preparation phase of the exam, and we also did not have enough information on how to conduct standardisation at the end of the exam. So far, we have been feeling ourselves as lay scholars.

One testing member in another university explained further:

Testing de çalışan arkadaşlara eğitim aldırarak istiyoruz, ama alanda ülkemizde yetkin insan sayısı da az. Yani bu konuda çalışma yapmış olan kalifiye insan sayısı da az diyebiliriz. Olan kişiler var tabii ki ancak sistemli bir eğitim aldırıyoruz, ancak yayınevleri aracılığıyla, o da niteliği düşük olabiliyor. Testing üyelerimizin daha iyi iş

çıkartabilmeleri için bu konuda bilgisi olan, teknik açıdan donanımı olan insanlardan bilgi paylaşımı yapılırsa çok iyi olur.

We want to train our members working at the testing unit, but the number of competent people in the field is low in our country. In other words, we can say that the number of qualified people who have worked on this subject is low. There are qualified people, of course, but we can't reach them to provide a systematic education, but through publishing companies, which can also be of low quality. In order for our testing members to do a better job, it would be great to share information from people who are technically equipped and knowledgeable about this subject.

The challenges related to secretarial work (N=311) comprise the third most frequently used type in this subcategory in the interviews. Most of the TUMs complained to have wasted their time on simple tasks such as carrying exam packs, copying, grouping, stapling exam papers, and preparation of proctor lists which could be conducted by an officer. A testing member, for instance, considered:

Aslında bilgisayarda sınavı hazırlamak açıkçası çok fazla bizi zorlayan bir mesele değil. Bunun zarflanması, basılması ki bu fiziki şartlar içerisinde yapılması çok zor. Sekreteryada düzeyindeki işler hem enerjimizi öldürüyor hem de motivasyonumuzu etkiliyor. Mesela soruları hazırladıktan sonra proof ların kontrolünü yapmak bile olay, dönüşler de uzun bir zaman alınca takibini bile yapamıyoruz. Ekstra işler çok, o yüzden zaman yetmiyor. Zaman yok, öyle bir zamanımız hiç olmadı oturup soru yazalım. Spesificationları hazırlıyoruz bir de word de tablolarıyla uğraşıyoruz. Şimdi ben onları mı yapayım, soru mu yazayım, yoksa hazır olan kısmın kağıtlarını mı ayırıştırayım. Cevap c şıkkı hepsi. Enerjimizi öldüren çoğu iş, bir sekreteryaya işi baktığınızda aslında.

In fact, preparing the exam on the computer is obviously not an issue that forces us too much. It is very difficult to envelop and print under these physical conditions. Secretarial work both kills our energy and affects our motivation. For example, after writing the test items, even checking the proofs is an issue, as the returns take a long time, so we cannot even follow up. There is a lot of extra work, so time is not enough. There is no time, we never had such a time to sit down and write a question item. We prepare specifications and also deal with making tables in word. Now, should I do them, write questions, or parse the papers of the ready part? The answer is option c, all of them. Most jobs that kill our energy are actually secretarial work.

A testing member added some more details and elaborated as follows:

Zimba dahil biz yapıyoruz. Ben hep şuna üzülüyorum; bu işi küçümsemek için değil ama bizim sınavı geliştirmeye yönelik hazır tecrübemiz varken 2-3 saatimizi, dediğim gibi gerçekten işi küçümsemek için değil ama bir başka kişinin yapabileceği bir işe, deneyim

gerektirmeyen bir işe, ayırmak bizi üzüyor. Ayrıca siz o işi bitireyim diye uğraştığınızda o verdiğiniz emek yorgunluk olarak size dönüyor. Bu işin bir başka boyutu. Geçenlerde bir hoca verilecekti birimize bu işleri yapmak için. Diğer hocalar fazla ders geleceği için sınav güvenliğini tehlikeye sokar diye itiraz ettiler. Bu soruları zaten proof aşamasında syllabus görüyor, hocalar görüyor. Ama özellikle birimin yasal zemini olmuş olsa, görev ve sorumlulukları belli olmuş olsa, o kişi de sizin biriminizde çalışan bir kişi, sekreter düzeyinde ya da asistan düzeyinde bir kişi olmuş olsa, sorumluluğu tanımlanmış biri olsa bu sorunları yaşamazdık, böylelikle enerjimizi daha kaliteli sorulara ayırırdık.

We do everything, including staples. I always regret this; I do not want to undermine this work, but when we have ready experience in improving the exam, it makes us sad to devote 2-3 hours to a job that does not require experience, like I said I really do not want to undermine it, but to the work that another person can do. In addition, when you try to finish that work, the effort you put in returns to you as fatigue. This is another dimension of the business. Recently, a teacher was going to be assigned to our unit to do these things. Other instructors who were afraid to get more lessons objected that it would endanger exam safety. The syllabus and the instructors see these questions already at the proof stage. But especially if the legal basis of the unit were clear, if the duties and responsibilities were clear, if we had a secretarial person working in our unit, as a secretary or an assistant, with a defined responsibility, we would not face these problems, so we would devote our energies to higher quality questions.

A testing member shared similar views but also put emphasis on time constraints due to secretarial work:

Sınav hazırlamak hammaliye işi. Fotokopi çekmeye gidiyoruz koridorun sonuna. Oradan oraya taşıyoruz kolilerde. Sürekli fıtığa davet, ağır şeyleri indir kaldır. Sürecimiz çok yoğun. N... hocanın da söylediği gibi, zaman konusu da belki bu yüzden, bu işler okulda yapılması gerektiği için sınav hazırlaması eve kalıyor, akademik işleri daha çok evde yapmamız gerekiyor. Burada daha fiziksel çalışıyoruz da asıl görevimizdeki işler eve kalıyormuş gibi. İş yükünün eve taşınması motivasyonumuzu etkiliyor, çünkü ekstra yorgunluk sebebi oluyor. Bunlara vakit ayırmak yüzünden evdeki zamanımı da ayırıyorum. Örneğin iki hafta önce sınav kağıtlarının çoğaltılması için uğraşırken kocaman bir makinemiz var, komple akıllı bir şey bozuldu. Yapmak için uğraştım, Teknik eleman çağırdık, gelmedi, farklı bir yerde çekelim dedik, izin alamadık. Ben uğraştım tamiriyle YouTube'dan izleye izleye yaptım. Bunun teknik bilgisi bilmek zorunda mıyız tartışılır.

Preparing the exam is like porter business. We're going to the end of the corridor to make copies. We carry the papers from place to place in boxes. Constantly we are at risk of herniation, carrying heavy boxes. Our process is very busy. As N... said, maybe that's why time is a challenge. Since these things have to be done at school, exam preparation is left at home, we have to do academic work more at home. We work more physically here, but it's

like our main job is staying home. Doing our workload at home affects our motivation because it causes extra fatigue. I also spend my time at home doing these. For example, we have a huge machine, something completely smart. Two weeks ago, while we were trying to copy exam papers, it broke down. I tried to fix it, we called the technical staff, who did not come. We decided to copy in a different place, we could not get permission. I dealt with the repair, I did it by watching it on YouTube. It is debatable whether we have to know the technical knowledge.

As most testing members stated, a member working in a testing unit should cope with stress (N=250). It was stated that the critical importance of exams in students' success and their transition to the department would draw attention to the exams and those who prepare the exams. This situation may cause reactions from different stakeholders. Most complained that reactions, pressures, time factors, and limited sources would bring stress. It is surprising that test preparation is considered to be rather stressful with no financial or emotional return. A testing member explained how she got stressed out in the office elaborating:

Bizim işler çok eleştiriye açık, hani bizim toplumu da biliyorsunuz eleştirinin bazen fazla dozunu kaçırıyoruz. Geri bildirim süreci oluşturduk hocalar eleştiri süreci olarak algıladı. Karşılığı olmadan emek veriliyor ama o emeğin karşılığı her zaman maalesef stres. Dolayısıyla kimse istemiyor, testing hocaların gönüllü olduğu bir alan olmuyor. Ben onu şeye benzetiyorum hocam. Futbolda kaleci vardır. 89 dakika kurtarır 90'da gol yer, okların hedefi olur. İşte öyleyiz maç boyunca stres altındayız acaba ne zaman gol yeriz diye. Maç da bitmiyor ki, bir pozisyon bitiyor diğeri başlıyor. Geçen ay A... ameliyatı oldu. Doktor stres yüzünden dedi. O izin alınca bizim de stres seviyemiz arttı, bakalım biz ne zaman ameliyat oluruz.

Our work is very open to criticism, you know our society, sometimes we overcriticize. We started a feedback process, yet the instructors perceived it as a chance for criticism. We put in our effort without compensation, but unfortunately, the reward for that effort always turns into stress. Therefore, no one wants to work, testing is not an area where instructors volunteer. I think it is football in which there is a goalkeeper. He saves the shoots for 89 minutes but concedes a goal at 90, then he becomes the target of arrows. Here we are, we are under stress throughout the match, wondering when we will concede a goal. The match does not end either, one position ends but another begins. A... had an operation last month. The doctor said it was because of stress. When she took leave, our stress level increased. We are wondering when we will have an operation.

A testing member shared the same perception and added some more details regarding pressure:

Zaman açısından bir pressure var, öğrenciler, hocalar açısından bir pressure var, yönetim açısından bir pressure. Bir de içsel pressureımız var. Kısa sürede, hızlı zamanda, az personelle kaliteli bir şey çıkarmak çok stresli bir iş. Diğer işleri düşünüyorum, özel bir yerde stres altında çalışıyorsun, mesela stateler çok yüksek yani ama başardıkları zamanda karşılığında hem maddi olarak hem de manevi olarak bir başarı hissi var. En azından o durum tatmin ediyor seni. Bizde o denge yok. İşin stresi yüksek ama sonucuna baktığımızda memnuniyet duygusu yok. Bunu anlatırken bile stresten dolayı tam açıklayamadım ben, heyecandan.

There is pressure in terms of time, students, instructors, and in terms of management. In addition, we have internal pressure. It is a very stressful job to produce a quality exam in a short time, passing fast, with few members. I think of other jobs, where you work under stress in a private place, for example. The states are very high, but when you succeed, there is a sense of achievement in return, both financially and morally. At least that satisfies you. We do not have that balance. The stress of the job is high, but when we look at the result, there is no sense of satisfaction. Even while telling this, I couldn't explain it fully due to stress, and excitement.

An important number of interviewees indicated that the source of the challenges encountered in testing practices results from the need to stay in the office (N=135). TUMs explained that they had to work in the office until five or six although the instructors go to their home after their classes are over. A more interesting point is that in the summer, TUMs are forced to prepare exams for the following year in the office even though the instructors take a vacation. The following extracts explain their feelings:

Ofis saatlerimiz var ve tüm gün buradayız. Bu ofisler şey gibi değil teach and go gibi değil. Hocalarımız sadece derse girip çıkıyor, yani tek ve asıl görevleri ders saati oluyor. Biz hem derse giriyoruz, ders sonrası da burada bulunmak zorundayız. Hoca dersini verip gidiyor, biz mesai kavramına takılıyoruz. Bizim asıl görevimiz hangisi? Birkaç saat az derse girdiğimiz için ek bir görev olduğunda ofis çalışanları yapsın deniliyor. Mesela hocalar 13:30 da dersi bittiğinde evine gidebiliyor. Bu dönem ben hiç 5'ten önce çıktığımı hatırlamıyorum, en erken 17:00 de çıkabiliyorum. Artı evde çalışmalarımız var.

We have office hours and we are here all day. These offices are not like teach and go. Our instructors only teach and leave the class, so their only and main task is their course time. We both lecture and have to work here after our lessons. The instructors give their class and leave, yet we get stuck with the concept of shift. What is our main task? Since we have only

a few hours of class reduction, when there is an additional task, we as office members are told to do it. For example, our instructors can go home this term when their lessons are over at 13:30. I don't remember ever leaving here before 5 pm this term, I can go at 17:00 at the earliest. Plus, we have work at home.

Tüm gün buradayız ve hocalar kağıt bırakacak, proof larını getirecek diye burası hep açık ve sürekli hocaları bekliyoruz. Bir memur gibi çalışıyoruz, gerçi daha kötü, çünkü onlar 12:00'de öğle yemeğine gidiyorlar, 13:00'de geliyorlar. Biz onu da yapamıyoruz. Hocaların ders saatine de biz dikkat ediyoruz, ilginçtir kendileri bize randevu veriyorlar. Burası hiçbir zaman boş kalmamalı diye uyarı alıyoruz, öyle bir şey var. Bizim hocalarımız ve memurlarımız öğlen havuza gidiyor ama biz gidemiyoruz. Mesela bir saat buranın boş kalması demek büyük intihar, hoca gelecek de burada kimseyi bulamayacak. Dün ben mesela tek kalmıştım tuvalete gittim, hoca gelmiş bizi bulamamış, "test ofis kapalı" diye şikayet etmiş. Uyarı aldık. Hocalar sabah, öğleden sonra, ders aralarında geliyorlar, ne zaman gelecekleri bile belli değil.

We are here all day; this office is always open and we are always waiting for the instructors to leave their papers and proofs. We work like an officer, though it is worse because they go to lunch at 12:00 and come back at 13:00. We cannot do that either. We also pay attention to the lecture hours. Interestingly enough, they make appointments for us. We are warned that this place should never be empty, well there is such a thing. Our instructors and officers go to the restaurant in the afternoon, but we cannot. For example, leaving this place empty for one hour is a big suicide, considering the instructors will come and will not find anyone here. Yesterday, for example, I was alone, and went to the toilet, an instructor came and could not find us, he complained that "the test office is closed." We got a warning. The instructors come in the morning, in the afternoon, and during the breaks, it is not even known when they will come.

A testing member explained office time in the summer:

Yazın burada hocaları göremezsiniz, ancak biz izin almadığımız sürece burada olmamız gerekiyor, ayrıca izin alamayacağımız kesin tarihler oluyor, testing olarak sınavlardan dolayı. Sadece yaz tatili değil sömestr tatillerimiz de öyle. Bunun karşılığı yok tabiki. Biz yazın da çok çalışıyoruz, hem revizyona gidiyoruz hem tekrar yaz okulu sınavı, sonra tekrar Eylül prof sınavı gibi. Standardizasyon toplantıları oluyor. Diğer ofislerden dönüt alıyoruz, onları inceliyoruz. Şubat tatilinde tekrar sınav yazıyoruz çünkü şubat prof sınavı geliyor falan. Yani hiç durmuyoruz. Bütünleme sınavları var, mesela resitler var. Daha kötüsü ise bunları ofisten yapmamız bekleniyor, evden yapayım dedim, yok denildi.

You can't see the instructors here in the summer, but we have to be here unless we get permission, there are also certain dates when we as testing members can't get permission due to exams. It is not only during the summer holidays but also during our semester

holidays. Of course, we do not have compensation. We are working hard even in the summer, going through revisions, preparing exams for summer school, then the prof exam in September. There are standardisation meetings. We get feedback from other offices and examine them. Then we are writing test items during the February break because the February prof exam is coming. That's why we never stop. There are make-up exams, and resits. What is worse, we are expected to do these from the office. I offered to work from home, but I was told no.

Furthermore, the need for isolation (N=97) is a noticeable concern of testing members. Preparing an exam, including the writing items and distractors, requires extra concentration and attention. However, most emphasized that they were distracted by instructors and students who come to ask about simple matters. One testing member expressed her feelings:

Bizim normalde derse giren hocalarla bu kadar iletişimimiz olması gerekmiyor. Biraz daha soyutlanmış olarak çalışmalıyız aslında. Sınav hazırlaması kolay bir iş değil, birçok parametreyi göz önüne almamız gerekebiliyor. Yoğun dikkat gerektiren bir iş yapmaya çalışıyoruz. Ancak sınav süreçlerinde çok fazla dikkatimiz dağılabiliyor ve bu durum bizim çalışmalarımızı sekteye uğratabiliyor. Mesela dün tam çeldiricileri yazarken bir öğrenci geldi, o gitti bir hoca gelip kağıtları bırakacaktım dedi. Teslim etmesi ve hocaya anlatması derken bir sürü zaman gitti. Daha önemlisi ise konsantrasyonumuz tamamen dağıldı.

Normally, we do not need to have so much communication with the instructors. Actually, we should work in a more isolated way. Exam preparation is not an easy task, we may need to consider many parameters. This work requires intense attention. However, we may have been distracted too much during the exam processes, which can violate our studies. For instance, yesterday while I was writing the distractors, a student came, after whom, a instructor came and said she was going to leave the papers. It took a lot of time to hand in the papers and inform her. But more importantly, our concentration was fully broken.

Another point made by the participants was their concern about teaching and test preparation at the same time (N=80). Teaching and test preparation have different responsibilities. Considering the instructors whose responsibilities only include the delivery of the course complain about their busy schedule, it makes the situation of TUMs more complicated since they do both teaching and test preparation at the same time. In the interviews, a number of directors complained that they could not decrease the number of teaching hours of TUMs due to legal obligations although they were

aware of their workload. As highlighted by a testing member, teaching and test preparation at the same time increase their workload and stress:

Bir yandan neredeyse hocalar kadar derse girmek, diğer yandan sınav hazırlamak canımızı çıkarıyor. Tek koltuk altında iki karpuz taşınamayacağı gibi imkansız kovalıyoruz. Aşırı bir yoğunluk olduğunu müdürümüz kabul ediyor ancak o da mecbur olduklarını, derslerimizi mümkün olduğunca reduce ettiklerini belirtiyor. Mesela benim ilk dönem 14 saat gibiydi, bu dönem 12 saat giriyorum. Ayrıca haftada dört saat ofis saatlerimiz oluyor. Öğrenciyle birebir iletişime girdiğimiz. 12 saat artı dört saat. Geçen sene maksimumdu 18 saat derse girdik ve sınav biriminde çalıştık. Belki diğer üniversiteler bizden az veya fazla giriyordur ancak aynı anda iki görev inanın zor. Bu yüzden kendi personal time dan bayağı bir vakit harcıyor oluyoruz sınavlara.

On one hand, we have courses, almost as much as the instructors; on the other hand, we prepare exams, which exhausts us. As we cannot carry two watermelons under one armpit, we are chasing the impossible. Our director admitted that we have an excessive intensity, but he stated that they have to give courses and that they have reduced our lessons as much as possible. For example, I had 14 hours in the first semester, but this semester I have 12 hours. We also have office hours like four hours a week. We have one-on-one contact with the students. 12 hours plus four hours. Last year the maximum was 18 hours; meanwhile, we worked in the testing unit. Maybe the members at other universities have more or less, but believe me, it is difficult to do the two tasks at the same time. That's why we spend a lot of our own personal time on exams.

Another director explained further:

Testingden beklenti çok fazla. Mevcut yasal zeminden dolayı hem derse girmeleri hem de işte bu kadar çok sınav yazmaları ve diğer süreçlerini organize etmeleri bekleniyor. Bu kadar iş yükünün altından kalkabilmeleri kolay değil. Müdür olarak testing member ların ders yüklerini düşürmek istiyoruz, ancak devlet üniversitelerinde pek mümkün olmuyor. Mesela, ilk dönem bir yandan testing birimindeki çalışmalarına ek olarak 20-24-28 arasında derse giriyorlardı, bu dönem 12 saate indi.

Expectations from testing are too high. Due to the current legal ground, they are expected to both give lectures and write so many exams, organize other processes. It is not easy for them to cope with such a workload. As the directors, we want to reduce the course load of testing members, but it is not possible in state universities. For instance, in the first semester, in addition to their work in the testing unit, they were taking classes around 20-24-28, it was reduced to 12 hours this semester.

Acknowledging there is always a change in the system, the participants approached the issue from the perspective of adaptation to change (N=66). It was

reported that constant changes would cause serious challenges for TUMs trying to produce something with their individual efforts in the system vacuum. A testing member, for instance, explained the issue as follows:

Sürekli değişim halindeyiz, tabi yeni sisteme adapte olmak bizi yoruyor. Kitap değişiyor, bu nedenle soru formatını değiştiriyoruz. Mesela geçen yıl final sınavını kaldırdık, quizlerin içeriğini değiştirdik, üstüne bu yıl da placement sınavıyla proficiency sınavını birleştirdik. Sürekli değişen duruma ayak uydurmak zorunda kalıyoruz. Çoğu üniversitede arkadaşımız da değişime adapte olmaktan şikayetçi.

We are in a constant state of change, of course, and adapting to the new system tires us. The book is changing, so we are changing the question format. For example, last year we abolished the final exam, changed the content of the quizzes, and this year we combined the placement exam with the proficiency exam. We have to keep up with the ever-changing system. Many of our friends at other universities also complain about adaptation to change.

Out of eight challenges related with Testing Unit Members, emotional factors such as demotivation, burnout, and anxiety were top issues. Working hard with limited human resources and dealing with stressful processes seemed to break their motivation. Besides, most TUMs criticised their need for professional training and doing secretarial work instead of designing language tests. Another major issue was the pressure from the administrators, their colleagues, and students, which was associated with stress at work. In the next section, the challenges related with instructors will be explained.

4.1.2.3. Challenges related with instructors

This subcategory includes eight emerging codes which are shown in Table 4.9.

Table 4.9. Challenges related with instructors

Emerging Codes	Number of Codes	%
Tense reactions	186	23,91
Distraction	142	18,25
Irresponsibility	123	15,81
Lack of empathy	88	11,31
No work discipline	66	8,48
Technology incompetence	65	8,35
Over-reliance on curriculum	58	7,46
Too many instructors	50	6,43
Total	778	100

The challenges related to tense reactions of the instructors (N=186) comprise the most frequently used type in this subcategory in the interviews. A testing member elaborated on clash situations and stated that:

Büyük ihtimalle özel üniversitede çalışanlar bu durumu size çok anlatmamıştır ama devlet üniversitesinde çalışanlardan çok duymuşsunuzdur. Bazı hocalarımız yapılan işe, emeğe hiç saygı göstermiyorlar ve nasıl diyeyim böyle üslupsuz bir şekilde, eleştirinin dozunu kaçırarak yorum yapıyorlar sınavlara. Tepkilerini bazen doğrudan bize gösteriyorlar, bazen de arkamızdan konuşuyorlar, şikayet ediyorlar, çok hoş olmuyor. Mesela daha dün bir native hocamız gelip “This exam is a disaster” deyip konuya girdi. Allahtan o sınavı hazırlayan arkadaşımız burada yoktu. Bu hoş bir durum değil, tabii ki bu durumdan dolayı motivasyon açısından çok sorun yaşıyoruz. Bizden önceki sınav biriminde çalışanların bu konuda daha kötü tecrübeleri olduğunu duymuştuk. Bazen biz de duygusal olabiliriz, etkilenerek cevap versek, iş büyüyecek. Maddi herhangi bir ekstrası yok bu işin, ama sürekli çalışmak zorundayız ve bunun karşılığında da sürekli clash halindeyiz, yani çatışma halinde. Manevi motivasyonumuzda baltalanıyor sürekli eleştiriye de maruz kalarak.

Most likely, those working at a private university did not tell you about this situation, but you have heard it a lot from those working at a public university. Some of our teachers do not show any respect for our work and effort, and they comment on the exams disrespectfully, by overcriticizing. Sometimes they show their reactions directly to us, sometimes they talk behind, they complain, which is not very appropriate. For example, yesterday, a native instructor came and got into the subject by saying “This exam is a disaster.” It is fortunate our friend who prepared that exam was not here. This is not a pleasant situation, of course. We have a lot of problems in terms of motivation because of this situation. We heard that those who worked in the exam unit before us had worse experiences. Sometimes we can be emotional too, if we respond with influence, the issue will get worse. We do not have extra income in this job, but we have to work constantly and in return, we are in a constant clash, that is, conflict. Our moral motivation is undermined by being constantly criticised.

A testing member added that almost all of the criticisms were negative and explained:

Acımasızca yorum yapanların ya da işte daha iyi bildiğini iddia edenlerin birçoğu testing de çalışmamıştır. Her defasında söylüyorum ben yerimi bırakmaya hazırım. Öyle olduğunu, daha iyi yapabileceğini düşünüyorsa bunu bize bir gösterebilir. Ancak bir dayanakla gelmiyorlar, açıklama yapamıyorlar. Sadece ‘bu sınav çok kolaydı’ diyorlar, hangi kısım kolaydı ya da neye göre kolaydı bir açıklama yok. Bırakın eleştiriye feedbackler bile hep olumsuz. Geçen dönem sınav sonrası toplayalım dedik, feedback varsa yazmalarını istedik, olumlu bir şey yok hocaların gözünde sadece olumsuz yorumlar var. Aslında Türkiye’de

genelde biliyorsunuz olumlu eleştiri pek tercih edilmez. “Güzel bir şey yaptım” duygusunu size yaşatacak bir dönüt alamıyorsunuz, bırakın appreciation almayı. Hani mecazi anlamda, pencere açık, pencereden de gelebilirler şikayete.

Most of those who comment harshly or claim to know better have not worked in testing either. I tell them I am ready to leave my place. If they think they can do better, they had better show us. However, they do not come with a justification, they cannot explain. They just say 'this exam was very easy', yet there is no explanation of which part was easy or according to what. Let alone the criticism, even the feedback is negative. Last term, we decided to collect feedback after exams, and we asked the instructors to write their feedback, there was nothing positive, only negative comments in the consideration of the instructors. In fact, in Türkiye, as you know, positive criticism is not favoured much. You can't get feedback that will make you feel "I did something nice," let alone get appreciation. In a figurative sense, the window is open, they can also come to complain from the window.

Distraction (N=142) has the second highest frequency among the challenges in this subcategory testing members face when codes are computed. A testing member explained as follows:

Yaptığımız iş, motivasyon ve konsantrasyon gerektiren bir iş. Basit bir şey için bile olsa, bir kişinin bu odaya gelmesi, bakması, bir şey sorup gitmesi bizim konsantrasyonumuzu bitirebiliyor. O ufak bir zaman belki benim yanlış yapmama sebep olabiliyor. Yanlış da fark etmesi geç olduğunda... Özellikle belli anlarda, mesela örnek veriyorum ben orada shuffle yaparken birinin gelip bana soru sorması, bütün konsantrasyonumu, dikkatimi, bütün motivasyonumu dağıtıyor açıkçası. Maalesef kağıt al ver, gidip gelmeler, hocaların gereksiz soruları ya da clash biçiminde tepkiler olabiliyor. Bazen gelen sorular da aslında onlara söylenen ya da toplantı da söylenen ya da handbookta olan bir şey bile aslında gelip soru mevzusu olabiliyor. Kibarca bile bu durumu dile getirdiğimizde “ne olacak ki, geldim, sordum, gidiyorum, çok zamanını almadım” diyor hoca. Biraz daha soyut kalabiliriz. Hocalar geldiğinde en ufak şeyde ‘üff’ yapıyor “ne biçim sormuşsunuz” diyor, o anda bizim motivasyonumuz gitti. Çünkü sınav işi full içine odaklanmanız gereken bir durum ve en ufak bir olayda çok çabuk konsantrasyon dağılıyor.

Our work requires motivation and concentration. A person coming to this room even for a simple thing, looking around, asking something can end our concentration. That little time might cause me to make mistakes. It may be too late to realize what is wrong... Especially at certain moments, for example, someone asks me a question while I shuffle, frankly, it distracts all my concentration, attention, and motivation. Unfortunately, they come for paper exchanges, curiosity, unnecessary questions, or reactions like clashes. Sometimes, they come here to ask what is told them at the meeting or what exists in the handbook. Even when we politely expressed this situation, the instructor responded, “What happened,

I came, asked, and I'm going, I didn't take much of your time.” We should stay more isolated. Some instructors sounded “uff” at minor issues and shouted "What kind of question did you ask?" At that moment, our motivation was gone. Because you need to focus on exam work fully, concentration may be lost quickly in the slightest event.

A testing member shared similar views but also put emphasis on the security issues:

Önemli bir nokta da güvenlik hususunda. Hocalar alakasız bir şey sormaya çok fazla girip çıkıyor buraya, bence bu çok büyük sıkıntı. Çoğu zaman öğrenci de gelebiliyor. Bizim kapımızın üstünde test ofis yazıyor. Akreditasyon dolayısıyla konuldu, birimlerin isimlerinin yazılması gerektiği söylendi. Ama onun orada olması, ilgiyi çekiyor, hocalar, öğrenciler merak ediyor ne yaptığımızı içeride. Sürekli birilerinin bu odaya gelmesi hem bizim dikkatimizi dağıtıyor hem de sınav güvenliği açısından dezavantaj.

Another important point is security. Instructors come here a lot to ask something irrelevant, I think this is a big problem. Most of the time, even students can come. There is a sign of ‘test office’ on our door. It was put due to accreditation, we were told the names of the units had to be written. But hanging the name of the office attracts attention so that instructors and students wonder what we are doing inside. The constant visit of others both distracts us and is a disadvantage for the safety of exams.

The third most frequent anxiety in this subcategory was testing members' concerns about the irresponsibility of instructors (N=123). Two testing members elaborated on how their irresponsibility influenced in the following extracts:

Biz sınavın güvenli bir şekilde uygulanması için uğraşırken bazen hocalar yönergeyi iyi okumuş olmayabiliyorlar. Hocalar yazılı instruction ları eksik okuyorlar. Öyle bir kültürümüz var genelde, hep “sözel instruction” almaya alışmışız. Yazılı şeyleri almakta zorlanıyoruz. Geçen hafta hoca sabahki sınav ile öğleden sonraki sınavın süresini örneğin karıştırmış, 15 dakika erken kağıtları toplamış.

While we are trying hard to conduct exams safely, the instructors sometimes may not read the instructions well. They read the written instructions incompletely. We have such a culture in general, we are used to receiving verbal instructions. We have a hard time reading the written ones. For example, last week, the instructor confused the duration of the morning exam with the one in the afternoon and collected the papers 15 minutes earlier.

2017-2018 ikinci dönem, bir hocamız konunun gerisinde kalmış, bize de o konuyu önemsiz gördüğü için söylememiş. Maalesef o konulardan biz cloze test yaptık ama üç net soru çıktı. O sınıfın sınavını tekrar yapmak zorunda kaldık. Bize erken söyleseydi, hızlı bir şekilde sınava müdahale ederdik.

In the second semester of 2017-2018, one of our professors lagged behind the subject and did not tell us because he considered it unimportant. Unfortunately, from those topics, we made a cloze test, but three clear questions came out. We had to retake the exam for that class. If he had told us earlier, we would have intervened in the exam quickly.

Many testing members indicated in the interviews that they were concerned about the lack of instructors' empathy (N=88). A testing member, for instance, put an emphasis on the lack of empathy and stated:

Eleştirirken bizim yerimize kendilerini koymuyorlar, basit düşünüyorlar sınıflarının başarısı düşük çıkmış diye sınava suç buluyorlar. Şikayet etmek için açıklarımızı bulmaya çalışıyorlar. Bazen yemeğe gittiğimizde bile insanlar kapıya gelmiş de bizi bulamamış diye şikayet ediyorlar. Öğleden sonra hocanın dersi yok ama eve erken gitmek için diyor ki öğlen gidip bakayım, 13:00'de gelmek yerine 12:30'da geliyor. Biz yemek için çıkıyorduk diyoruz kibarca, beş dakika almaz diyor ama öğlen alt üst oluyor. Sadece kendi açısından düşünüyor. Dün benim profrederlarımından bir tanesi özellikle teneffüste, aralarda gelip bakıyor. Neden mi ders bitince özel dersine gidecek. Biz sınava mı konsantreyiz yoksa kağıtlarla mı uğraşıyoruz, bizi düşünen yok. İstisnasız hepsi bunu yapabiliyor yani. Uygun üslubu bilen kişi sayısı az. En azından emeğimize değer vermeleri ve bizim açımızdan da olaylara bakmalarını bekliyoruz. İlla beğenmeleri gerekmiyor.

They do not put themselves in our shoes when criticizing, yet they think simply. They put the blame on the exam when the success of their class is low. They are trying to find our vulnerabilities to complain. Sometimes, even when we go to lunch, they complain that they come but cannot find us. Instructors may not have a class in the afternoon, but in order to go home early, they want to have a look at the papers at noon, coming here at 12:30 instead of 13:00. Two weeks ago, we politely said we were going out for lunch, he said it would not take even five minutes, but it ruined our lunchtime. He only thought for himself. For instance, yesterday, one of my proofreaders came and wanted to have a look during breaks. That is why she will go to her private course after the lesson is over. No one cares about us, whether we are concentrating on an exam or dealing with papers. All without exception can do this. Few people know the proper language. At least, we expect them to value our labour and look at things from our point of view. They don't necessarily have to like the exam.

Among the challenges, the low work discipline of instructors (N=66) was a great concern for testing members. A testing member cited her concern about their low work discipline:

Hocalarımızın iş disiplini neredeyse hiç yok denecek kadar az. Olanlar da özel dershanelerde ekstradan çalıştıklarından yeterince buraya özen göstermiyorlar. Mesela bazen sınıfta programa göre geri kalabiliyorlar, bize bile haber vermiyorlar. Biz itiraz

dilekçelerinden öğreniyoruz, öğrenci şikayet ederse. Bazı hocalar da öngörülemeyen bir şey çıktı diye haber veriyor ama sınavdan 1-2 gün önce yani son dakika sürprizi. Öğretim görevlisi arkadaşlarımız okuma, kağıtları teslim etme, notları grime gibi görevlerini geciktirebiliyorlar. Aslında görevlendirme konusunda bir yönergemiz var, ama yönergeye bile uymuyorlar. Checklist geliştirdik hocalar için takibi kolay olsun diye, o checklist bile yarım oluyor. Dün bir hoca kağıtları sıralanmamış bir şekilde getirdi, acelesi varmış, içinde signature listesi yoktu. Hocaların yapmaları gereken sorumluluklar ilgili sıkıntı yaşıyoruz ve onun muhatabı biz oluyoruz. Bir görev verildiğinde, “off bu ne ya bunu da mı biz yapacağız” gibi bazı hayıflanmaları oluyor. Bu durumlarda canımız sıkılıyor, o tarz şeyler oluyor.

The work discipline of our instructors is almost non-existent. The ones who have self-discipline work extra in private language schools and do not pay attention to here. For example, sometimes when they can fall behind programme in the classroom, they do not even inform us. We learn from the appeal forms if a student makes an objection. Some instructors also report 1-2 days before the exam that something unpredictable has happened, like a last-minute surprise. Our instructors can delay their tasks such as reading, handing over papers, and entering scores. We actually have a directive on assignment, but they don't even follow the directive. We have developed a checklist so that it gets easy for them to follow, even that checklist is half. Yesterday, a instructor brought the papers in an unordered way, she was in a hurry, and there was no signature list in it. We are having such trouble with their responsibilities and we should deal with them. When a task is given, they have some lamentations such as "off what is this or are we going to do this?." In these situations, we get bored.

Another point was mentioned by a testing member:

Sınav zamanları hastaneden rapor alan sayısı çok, izinli sayısı çok. Hadi hastaneye bir şey diyemiyorum, yönetim neden izin veriyor? Biliyorlar nasıl yoğun olduğumuzu. Biz hastalanmamaya çalışırken, nezle oldum diyene üç gün, beş gün rapor veriliyor. Geçen güz döneminde son gün dünya kadar rapor geldi. Notlar geç girildi. Çok büyük sıkıntıydı.

A number of people received a report from the hospital during the exam period. Leave alone the hospital, why does the management give permission? They know how busy we are. While we are trying not to get sick, those who have flu are given a three-day, five-day report. The student limit in the last fall was exceeded, and we realised a lot of sick reports were sent. Grades were entered late. It was a big problem.

In addition to the concerns stated above, technology incompetence of lectures (N=65) is a factor contributing to anxiety. A testing member cited her concern about their technology incompetence:

Not edin lütfen, hocaların teknoloji özürü olmasý da testing sürecine çok büyük etki ediyor. Ortaçağ skolastik dönemindeki gibi yeni bir şey yapmaya korkuyoruz, teknolojik anlamda. Bilgisayar kullanmaya ve USB ye geçtiğimizde evet bayağı bir direnen oldu, “ay ben nasıl yapacağım” şeklinde. Yönetimin de desteğı ile 4. ya da 5. sınavın sonunda öğrendiler. CD’ye alışmışlar ya, CD’den daha kolay olduğunu anlatamadık. Biz bile zor ikna ettik. Aya çıkmış gibiydik. Hemen her sınavda, çağırıp fişi takmamış hocalar USB, cd çalar çalışmıyor diye bizi çağırıyorlardı panik içinde. Teknolojiyi ve dijitalleşmeyi daha fazla kullanmalıyız.

Please note that the instructors are technologically incompetent, which greatly affects the testing process. We are afraid to do something new in a technological sense, as in the medieval scholastic period. There was a lot of resistance when we shifted to using a computer and USBs. They resisted “how am I going to do it?” With the support of the management, they started to use it at the end of the 4th or 5th exam. Having been used to CDs, they did not want to accept that it was easier than CDs. We had a really hard time convincing them. It was like we stepped on the moon. In almost every exam, teachers who did not plug in were calling us claiming the USBs, CD player did not work. We should use technology and digitalisation more.

Some of the testing members approached the challenges by emphasizing instructors’ over-reliance on the curriculum (N=58). One of the testing members specified her concern on this issue as follows:

Teachingimiz yetersiz, teachingimiz kötü, yani kötü derken skill öğretmiyoruz. Mesela note-taking sorusu soramıyoruz çünkü hocalarımız bir müfredat belirlemişler ve o müfredatta olmadığından note-taking öğretmiyorlar. Syllabusımız çok yoğun. Bazı yöneticilerin ve öğrencilerin gözünde hocalarımız müfredat öğretmeniyiz.

Our teaching is inadequate and bad, that is, we do not teach skills. For example, we cannot ask a note-taking question because our teachers have determined a curriculum and they do not teach note-taking because it is not in the curriculum. Our syllabus is highly intensive. From the point of some administrators and students, our instructors are like curriculum teachers.

Finally, testing members rarely mentioned the high number of instructors (N=50). As explained by the participants:

Hoca sayımız o kadar çok ki her hocaların da farklı özellikleri var. Sistem getirmeye çalışıyoruz ama gerekli desteğı bulamadık. Sistem de olmayınca her hoca ayrı telden çalabiliyor. Mesela bu durumda standardizasyon nasıl sağlanacak, hani biz burada 80 kişiyi standardize edemiyoruz.

We have so many instructors each of whom has different characteristics. We are trying to bring a system, but we could not find the necessary support. In the absence of a system, each instructor can play on a separate string. For example, how will standardisation be achieved in this case? We cannot standardize 80 people here.

The results revealed that the highest-ranking challenge was the instructors' tense reactions toward the TUMs. It was stated that preparing a test was itself an invitation to negative attitudes and harsh reactions at the slightest error. Another major challenge was the instructors' distraction which was supposed to break their concentration. The other most striking problems included irresponsibility, lack of empathy, and low work discipline of the instructors. These problems are followed by the instructors' technology incompetence and over-reliance on curriculum. The next subsection deals with challenges related with students.

4.1.2.4. *Challenges related with students*

This subcategory includes ten emerging codes which are shown in Table 4.10.

Table 4.10. *Challenges related with students*

Emerging Codes	Number of Codes	%
Too many students	170	15,47
Low English proficiency	165	15,01
Gap between levels	159	14,47
False level placement	132	12,01
Unwillingness	127	11,56
Reactions to tests	111	10,10
Lack of critical thinking	85	7,73
Internalisation of productive concepts	75	6,83
Lack of information in optic forms	43	3,91
Low cognitive-cultural level	32	2,91
Total	1.099	100

The most frequent anxiety in this subcategory was testing members' concerns about the high number of students (N=170). One testing member mentioned in the interview:

Özellikle çok kalabalık sınıflarda sınavlar için öğrencileri çekip yeni bir sınıf açmak zorunda kalıyoruz. Eğitim görürken bile bu çocuklar biraz alt alta üst üste ders alıyorlar. Küçük bir sınıfta 40 kişi sınava giriyor. Hoca şurada, öğrenci burada oturuyor, öğrencinin nefesi adeta hocanın yüzünde. Mesela bizim sene başında 1.200-1.300 civarında

öğrencimiz oluyor. Ne kadar büyük bir rakam. Öğrenci sayısı fazla olunca, bazı öğrenciler de sınavda kayboluyor. Birazcık temeli olanlar bir şekilde idame ediyor ama sıfırdan başlayanlar kayboluyorlar. Sonra da “3’te 1’i geçti, bu başarı mı?” diye sınava suç bulunuyor.

Especially in crowded classrooms, we have to separate students and open a new class for exams. The students sit one on one even while they are taking a lesson. 40 students take an exam in a small class. The instructor is sitting there while the student is sitting here, it is almost as if the student's breath is on the instructor's face. For instance, we have about 1,200-1,300 students at the beginning of the year. What a big number. When the number of students is high, some students get lost in the exam. Those with a bit of foundation somehow survive, but those who start from zero get lost. Questioning “Is it a success when one third passed the exam?”, they put the blame on exams.

A testing member explained in a more detailed way:

Öğrenci sayısından bahsederken şeyi de unutmamak lazım. Şimdi biz sadece burada hazırlık öğrencisine sınav hazırlamıyoruz. Mesela geçen sene 7000 kişilik Erasmus sınavını biz hazırladık. Bizim kendi öğrencilerimiz de 3000 civarı. Her hafta quizi, ara sınavı, dönem sonu sınavı. Erasmus sınavından hemen sonraki hafta vize vardı. Toplam sayımız 10.000 civarı. Arada başka bir şeyin sorumluluğunu aldığımızda sistem çöküyor yani. Buranın minimum sekiz kişi olması lazım. Dört kişi kesinlikle yürümüyor. Bu sistem belki beş sene önce ilk kurulduğunda X Üniversitesi'nin öğrenci sayısı daha azdı ve işler bir şekilde yürüyormuştu ama şu anda kesinlikle yürümüyor.

When talking about the number of students, one issue should not be forgotten. We are not just preparing exams for preparatory students. For example, we prepared the Erasmus exam for 7.000 people last year. Our own students were also around 3.000. There was a midterm the week after the Erasmus exam. Our total number was around 10.000. Thus, when we take responsibility for something else, the system crashes. There should be a minimum of eight people here as four people are definitely not enough. When this system was first established five years ago, the number of students at X University was less and things were working in a way, but now it is definitely not working.

The challenges related to low English proficiency of students (N=165) comprise the second most frequently used type in this subcategory in the interviews. A testing member, for instance, reported:

Öğrencilerin çok düşük farkındalıklarla ve yetersiz bilgiyle gelmesi büyük bir problem. Maalesef öğrencilerimizin çoğu A1 düzeyinde. Milli Eğitimde o kadar saat İngilizce görmelerine rağmen %50'sinden fazlası 0 geliyor. Mesela, sene başındaki muafiyet

sınavında 1.200 öğrencinin 700-800'ü 0 geliyor ve A1'e yerleştiriyoruz. Ne kadar ciddi bir rakam. Son iki yıldır B1 herhalde hiç yoktu.

It is a big problem that students come with very low awareness and insufficient knowledge. Unfortunately, most of our students are at A1 level. Even though they receive English for many hours during National Education, more than 50% come as 0. For instance, in the exemption exam at the beginning of the year, 700-800 of the 1.200 students performed at 0 level and we placed them in A1. What a serious number! For the last two years, we have not had B1 level students at all.

The gap between students' levels (N=159) has the third highest frequency among the challenges in this subcategory testing members face when codes are computed. A testing member elaborated on the issue as follows:

Yaptığımız placement sınavı ile seviyeleri ayırıyoruz ancak A1 ve A2 de öğrenciler arasındaki seviye farklılıkları çok fazla. Mesela sınıf içerisindeki dengeler açısından en çok fark, özel okuldan gelen öğrenciyle devlet okulundan gelen öğrenci arasında. Zaman zaman level'larda uygunluk da olmayabiliyor, dolayısıyla çok farklı profiller oluşuyor. Seviye aralıkları öğrenci farklılıkları açısından hiç yeterli değil. Örneğin, sene başında intermediate başlayan öğrencilerle elementaryden gelip intermediate olmuş öğrencilere aynı seviyede bir sınav veremiyoruz. Önceden bilgi birikimi ile gelen bir öğrenciyle burada Elementary sıfırdan başlayan bir öğrenci arasında bile çok fark var. İki gruba da aynı sınavı verirsek olay çıkıyor. En başta hocalar itiraz ediyor, sınav çok zordu öğrenciler için diye.

We separate the levels with the placement exam, but the level differences between students in A1 and A2 are very high. For example, the biggest difference in terms of balance in the classroom is between the students coming from private schools and the ones from public schools. From time to time, levels may not be suitable, so different profiles emerge. Level ranges are not sufficient in terms of student differences. For example, we cannot give the same exam to the students who started intermediate at the beginning of the year and the ones who came from the elementary level and became intermediate. There is a huge difference between the students who come with prior knowledge and the ones who start here from the elementary level like zero. If we give the same exam to both groups, reactions arise. At first, the instructors object claiming the exam is very difficult for the students.

Another member explained further:

Önemli bir sorun da filoloji grupla diğer bölümlerdeki öğrencileri ayrı gruplamıyoruz. Eskiden placement sınavımız daha belirleyiciydi, ancak placementın seviyesi düştüğü için sınıflar belirlendiğinde sınıf içindeki seviyeler arasındaki fark arttı. Bu iki grubu ayıralım

diye önerdik ancak olmadı. İngilizce öğretmenliği ve edebiyat bölümü ile mühendislik ve tıp öğrencilerini ayırabilsek en azından farkı azaltabiliriz.

Another important problem is we do not separate the philology group from the students in other departments. In the past, our placement exam was more decisive, but as the level of placement was lowered, the gap between levels within the class widened. We suggested separating these two groups, but we were not permitted to. If we can separate English language teaching and literature departments from engineering and medical students, we can at least reduce the difference.

Almost all testing members highlighted in the interviews that false level placement seems to lead to other challenges (N=132). Some testing members explained their concerns as follows:

Programa kabul şartları ile ilgili, öğrenciler şehir dışında olduğundan sınava girmiyorlar. Sınava girmeyince de random bir kura ile sınıflara atılıyorlar. Bir ikincisi de bilerek hazırlıkta kendi tabirleriyle “yatmaya” gelen öğrenciler var. Bu öğrenciler sınava girip, sırf hazırlıkta kalayım diye random bir puan alıyor. Bu durumda, B1 seviyesinde A1 çocuğu oturuyor, A1 de B1 oturuyor. Ve bence kur aralıkları da çok yeterli değil. Placement zaten olmayınca da proficiencynin de bir anlamı olmuyor.

Regarding the admission requirements to the programme, students do not take the proficiency exam because they are in another city. When they do not take the exam, they are scattered into the classrooms with a random draw. Besides, there are some students who want to “let the grass grow” in their terms, during the preparatory year. These students get a random score so that they can stay in the preparatory programme. In this case, an A1 student may be at level B1, and a B1 student may be at A1. And I think the levels are not very sufficient, either. Proficiency is meaningless when there is no placement anyway.

En büyük yaşadığımız sıkıntı öğrenciler placementı girmeyebiliyor ve sınava girmeyen öğrenciyi eğitim hakkından muaf tutamayacağımız için ya da onları sınava girmeye zorlamak için elimizde bir tool veya yönerge olmadığı için dönemin ilk birkaç haftası bu durum bize sıkıntı yaratıyor. Yani placementı girmemiş bir öğrenciye yeniden placement vermek zorunda kalıyoruz ya da onlar otomatik beginnera yerleşiyorlar. Bir kere bir bayram sonrasına denk gelmişti, sınava çok girmeyen olmuştu, o zaman komple sınavı tekrar yapmıştık. Genelde ilk hafta içerisinde hocalar bize kendi fark ettikleri öğrencileri gönderiyorlar ya da öğrenci “ben aslında P1 değilim ama giremedim” diyor. Hoca ona mesela bir sayfalık bir yazı yazdırıyor ve kanaatini de belirten bir dilekçe yazıyor bize.

The biggest problem we face is that students may not be able to take the placement, and since we cannot exempt the student who does not take the exam from the right to get an education, or because we do not have a tool or legislation to force them to take the exam, this situation creates problems for us in the first few weeks of the semester. Therefore, we

have to give a new placement to the students who have not taken the placement or those who are automatically placed at the beginner level. Once, it was after a religious holiday, there were many people who did not take the exam, then we repeated the whole exam. Generally, within the first week, the instructors send us the students they noticed, or the students may say "I am not actually P1, but I could not take it." Then, the teachers make him/her write a one-page article and write us a letter stating their opinion.

Another testing member added:

Çok klasik ve yaygın bir sorun, "ben sıfırdan başlayayım" düşüncesiyle seviyesini belli etmemek isteyen öğrenciler oluyor. Zaten zero beginnerım, hiçbir şey bilmiyorum düşüncesiyle de girmeyebiliyorlar placementa. Bir başka deyişle ben zaten beginner olacağım diyor çocuk hiç girmiyor sınava. Dolayısıyla biz tespit edemiyoruz umursamadığından mı girmedi yoksa ben zaten zero beginnerım diye düşündüğünden mi girmedi. Ama öğrenciler biliyorlar sınava girmezlerse beginner sınıfına yerleştirilecekler. Sınav öncesinde ve sınav esnasında bizim çok önemli olduğunu vurgulamamıza rağmen yine de umursamayıp sınava girmeyen öğrenciler oluyor. Sonra da sınıf değişikliği yapmak istiyorlar.

A very classic and common problem is that there are students who do not want to show their appropriate level with the thought of "I'll start from the beginning." They do not take the exam thinking "I am already a zero beginner, and do not know anything." In other words, the students do not take the exam because they believe they will be a beginner anyway. Therefore, we cannot determine the reason why they have not taken it whether because they have not cared, or because they have thought that they are already a zero beginner. But they are aware that if they do not take the exam, they will be placed at the beginner level. Although we emphasize before and during the exam that it is very important, there are still students who do not care and do not take the exam. Later, they want to change their class.

One testing member in another university explained further:

Öğrenciler muaftiyet sınavının zamanının farkına varmıyorlar. Yani muaftiyet sınavını kaçırmış oluyorlar. Kaçırdıkları zamanda 2. aşama olan level belirleme, düzey belirleme sınavına giriyorlar. Daha ilginç olanı ise düzey belirleme sınavı esnasında üç aşamamız olmasına rağmen bazı öğrenciler aşamalardan birine girmesinin yeterli olacağını düşünüyorlar. Bu durum ile ilgili açıklama yine internet üzerinde belirtiliyor. Ancak tek aşamaya girip çıkıyorlar ve dolayısıyla düşük leveldan, onların kendisini göstermeyen leveldan başlamak zorunda kalıyorlar. Böyle bir problemimiz var öğrenciler adına.

Students do not realize the date of the exemption exam. In other words, they miss the exemption exam. When they miss, they take our placement exam, which is the second step. What is more interesting, although we have three parts in the placement test, some students

think that it will be enough to take only one part. The explanation of this situation is also stated on the internet. Nevertheless, they take only one stage, and therefore they have to start from the low level, the level that is not appropriate for them. We have such a problem.

As a number of participants indicated in the interviews, students' unwillingness affects their reactions to the test prepared (N=127). One testing member expressed her concerns about their unwillingness:

Öğrencimiz çalışmıyor, gerçekten biz onları çalıştıramıyoruz. Öğrenci derste yatıyor, uyuyor. Hep kendi açımızdan baktık, kendimizi suçladık ama gerçekten karşımızda böyle bir profil var. Sınavdan haşat bir şekilde çıkmış hiçbir şey yapmak istemeyen “nasıl olsa geçerim” diyen bir öğrenci var karşımızda. Bu sene hep yatıp uyumak isteyen, o liseden çıkış dönemi rahatlama isteği, “o hazırlık yata yata geçersin” modu var herhalde. Motivasyonları yok. Bütün sınıfla ders yaptığınız bir ortam yok artık.

Our students do not study, we really cannot make them study. Some students sleep during the lesson. We have considered the issue from our own point of view and we blamed ourselves, but we really have such a profile before us. We have a student who comes in a mess out of the university entrance exam and does not want to do anything here saying “I will pass anyway.” This year, I guess there is a mood among the students that want to have one year off and to relax during the exit period from high school, and the belief they will pass without any effort. They are not motivated. There is no longer an environment where we can integrate the whole class.

A testing member added that they tended to cheat instead of studying and explained:

Öğrenciler o kadar isteksiz ki sınava çalışmak yerine sınav sırasında kopyaya teşebbüs edebiliyorlar ve bu ciddi oranda olabiliyor. Bazı öğrenciler o kadar rahat ki kimliklerini unutuyor, getirmiyorlar, kimlik bilgilerini yazmıyorlar. Öyle sıkıntılar oluyor.

The students are so reluctant that they may attempt to cheat during the exam instead of studying for the exam. This can be in serious amounts. Some students are so comfortable that they forget their IDs, they do not bring them or they forget to write their ID information. We have had such troubles.

Another testing member drew the attention to another side of the issue:

Bizim geçememe ile ilgili en büyük problemimiz şu, öğrenciler devamsızlık takibi yapamıyorlar ve kalabiliyorlar. Derse hiç gelmeyip veya derse gelip ancak hiç dersle ilgilenmeyip, beni geçirin diye yalvaranlar var. Hocalar da bıktı, dersi dinlemek isteyen öğrenci gelsin, dinlemek istemeyen gelmesin. Bu durum dersi etkiliyor, sınavları etkiliyor, bu yüzden sınavları daha kolay hazırlamak zorunda kalıyoruz.

Our biggest problem is that students cannot follow up on their absence and fail. There are those who never come to the class or have attended but never paid attention to the lesson and then beg us to help them pass. The teachers are also fed up, the students who want to listen to the lesson should come while those who are not interested in the lesson should not come. This situation affects the lesson and the exams as we are required to prepare easier exams.

Another major concern of the directors and testing members related to stakeholder-related challenges is students' reactions to tests (N=111). As stated by a testing member:

Öğrenciye uygulanan anketlerde biz hedef tahtası gibiyiz. Mesela biz sınav yapılırken koridorda gezdiğimiz için, 'siz niye koridorda geziyorsunuz, niye bir sınıfın başında değilsiniz?' diye soruyorlar ve bize CIA diyor öğrenciler. Onların algıları genelde nota göre zaten belirleniyor. Şu an istatistiksel olarak sınav notları genelde iyi olduğu için onlar açısından olumlu seyrediyor. Ama kuru tekrar eden öğrencilere sorsanız tek suçlu biziz. Sonuç odaklı bir toplum yetiştiriyor. İkinci yıl okuyan öğrenciler de herhalde konuşsa diyecekler ki sorunlu sistem.

We are like a target board in the questionnaires applied to the students. For example, because we walk in the blocks during the exam implementation, they ask why we are walking in the hallway and why we are not in a classroom. The students call us CIA agents. Their perceptions are usually determined by their scores. Currently, exam grades are statistically high, there is no problem for them. But if you ask the students who repeat the level, we are guilty only. A result-oriented society is growing. If they ask the second-year students about the issue, they will probably say the system is problematic.

A testing member shared the same perception and added some more details:

Öğrenciler genetik olarak itiraz etmeye meyilliler. İçsel olarak intrinsic bir şekilde ben niye bu notu aldım hocam daha yüksek bekliyordum şeklinde hep itiraz halindeler. Genelde argümanları sınavların ders içerikleriyle uyuşmadığını ve sınavların seviye olarak kitaba göre daha zor olduğunu söylüyorlar.

Students are genetically inclined to object. Intrinsically, they are always objecting to the exams asking for the reason why they get such a low score and are hoping to receive higher. In general, their arguments are that the exams do not match the course content and that the exams are more difficult in terms of level than the book.

As one of the challenges related to students was a lack of or limited critical thinking skills (N=85), the exams prepared seemed to be abstract for them. The following extract exemplifies:

Öğrencilerin kritik düşünme becerileri ve strateji üretme yetileri çok kötü. Ezberci bir mantıkla geliyorlar, sonra da hazırladığımız soruların zor olduğunu söylüyorlar. Reading sınavındaki bir metinde olay örgüsünü bile çıkaramaz durumdalar. Sınavlarda gramer odaklı veya test odaklı bir sınav hazırlamayacağız, dört beceriyi ölçen modern sınavlar hazırlamak için çalışmalar yapmıştık, ancak öğrencilerin kritik düşünme becerilerinin zayıflığından dolayı soruları eskisi gibi hazırlamaya başladık.

Students' critical thinking skills and their ability to produce strategies are very poor. They are rote learners, then claim that the exams we have prepared are difficult. They cannot even realise the plot of a text in a reading exam. We conducted some studies in order not to prepare grammar-oriented or test-oriented exams, but to prepare modern exams that are aimed to measure four skills. However, due to the weakness of the students' critical thinking skills, we started to prepare the exams as before.

Another challenge experienced by testing members was reported as internalisation of productive concepts (N=75). One testing member focused on the importance of productive skills:

Sınavlarımızdaki en önemli sorunlardan biri de beceri odaklı olmamaları. Özellikle vakıf üniversitelerinde öğrenci odaklı sistem yanlış anlaşılıyor. Öğrenciler productive skillerde sorun yaşadıklarından daha çok test soruları sormaktayız. Konuşma sınavını mesela kaldırdık. Yazma sınavını da sadece quizlerde yapmaya başladık. Productive skillerin proficiency sınavında olmaması dersleri de etkilemekte, çünkü hocalar bu becerilere daha az zaman ayırıyorlar. Öğrencilerimizin bir kısmının eğitimlerine önem vermeyip ÖSYM'nin yaptığı sınavlara yönelmesi de bizim sınavların scope unun değişmesine etki etmektedir.

One of the most important problems with our exams is that they are not skill-oriented. Especially in foundation universities, the student-oriented system is misunderstood. Since students have problems with productive skills, we included more test questions in our exams. For example, we removed the speaking part. Besides, we started to do the writing part only in quizzes. The absence of productive skills in proficiency exams also affects the courses as our instructors spend less time on productive skills. The fact that some of our students do not attach importance to their education and take the exams held by ÖSYM also affects the change in the scope of our exams.

As highlighted by the participants, it seems that a lack of information in optic forms causes testing unit members to waste their time and energy decreasing their motivation (N=43). A testing member explained the situation in the interview as follows:

Bazen sınav tutanaklarında eksik işler olabiliyor, yani öğrenciler optik doldurmayı unutuyorlar falan. Bazen sınav gözetmeni de eksik form getiriyor, sınava girmeyen öğrencinin numarasını yanlış kodluyor ya da kodlamıyor. Talimatlara %100 uyulmadığı zaman bu eksik bilgiler bizi zora sokabiliyor. Öğrencinin kağıdını bulmak vakit kaybettiriyor. Geçenlerde answer keyde bazı öğrenciler a yerine b yazmış. Aslında öyle çok küçük bir hata ama hangi öğrenci olduğunu bulmak o kadar kolay değil. Kitapçık işaretleme sorununa ek olarak ismini yazmayanlar oluyor.

Sometimes, there is some missing information in the forms and exam documents, that is, students forget to fill in optics. Sometimes instructors bring incomplete forms. Although they are required to note the number of students who do not take the exam, they mark incorrectly or do not mark. This incomplete information can put us in trouble when the instructions are not followed 100%. We finally find out which student it is, but it wastes our time. In the last exam, some students marked b instead of a in the answer key. Actually, it was such a small mistake, but it was not easy to find out which student it was. This happened 2-3 times in the last exam. In addition to the booklet marking problem, there are those who do not write their names.

One final reason why the participants worry seems to be the low cognitive and cultural level of students (N=32). One testing member explained how she felt:

Akademik anlamda, öğrencilerin kültürel seviyeleri o kadar düşük ki çok sıkıntı olabilir, bu sadece bizim okula has bir şey değil. Biz bir parça ya da bir soru hazırlarken, bunu bilirler mi diye düşünüyoruz sürekli. Genel olarak bir öğrenci profilindeki kültürel düşüklükten dolayı bir reading parçası veya bir cloze testteki konu seçimi bile çok zorlayıcı oluyor. Genel bir sorun ama sorun nihayetinde öğrencilerin.

Academically, the cultural level of the students is so low that it can be very distressing. This is not something unique to our school. While we are preparing a part or a question, we always consider whether they are familiar with it. In general, due to the culturally low level of students' profiles, even the choice of topic in a reading part or a cloze test is very challenging. It's a common issue, but a problem in the end.

A testing member drew the attention to another side of the issue:

Konu seçimindeki yaşanan problem, çocukların genel cognitive levellerindeki sorunlardan dolayı. YKS'yi geçen çocuk muhakeme yapamıyor, genel kültür seviyesi çok düşük, ilgi ve motivasyonları da olmayabiliyor. Magazin olmadıkça gündemdeki haberlere veya olaylara karşı duyarsızlar. Sınavlardaki seçenek seçme onların cognitive seviyelerini etkilemiş görünüyor.

The problem in choosing a topic is due to the problems in the general cognitive levels of the students. A student who passes the YKS cannot reason since his/her general culture

level is very low, and they may not have interest and motivation. Unless it is news from magazines, they are insensitive to current news or events. Choosing options in exams seems to have affected their cognitive levels.

According to the results of the study, the highest-ranking challenge in this category was the number of the students taking exams at EPPs. Most TUMs complained about the low level of the students in addition to the gap between their levels, false level placements, which was claimed to complicate the process. Another major challenge was the students' negative attitudes toward language tests. The other most striking problems included their lack of critical thinking and cognitive-cultural level, which was considered to influence their comprehension in exams. The next subsection deals with testing processes-related challenges.

4.1.3. Testing processes-related challenges

The third frequently stated type of challenges is testing processes-related challenges (N=3.949; 22,89%) encountered in testing practices at EPPs in higher education. This category consists of the challenges during the preparation, implementation, evaluation, and post-evaluation phases.

Table 4.11. *Testing processes-related challenges*

Sub-categories	Number of Codes	%
Challenges during preparation	945	23,93
Challenges during implementation	699	17,70
Challenges during evaluation	1.150	29,12
Challenges during post-evaluation	1.155	29,25
Total	3.949	100

As Table 4.11. presents, the highest-ranking subcategory belongs to challenges during preparation (N=945; 23,93%). The second highest-ranking subcategory is related to challenges during implementation (N=699; 17,70%). Challenges during evaluation (N=1.150; 29,12%) and post-evaluation (N=1.155; 29,25%) are in the last two sub-categories. The Figure 4.4. shows the sub-categories and their codes.

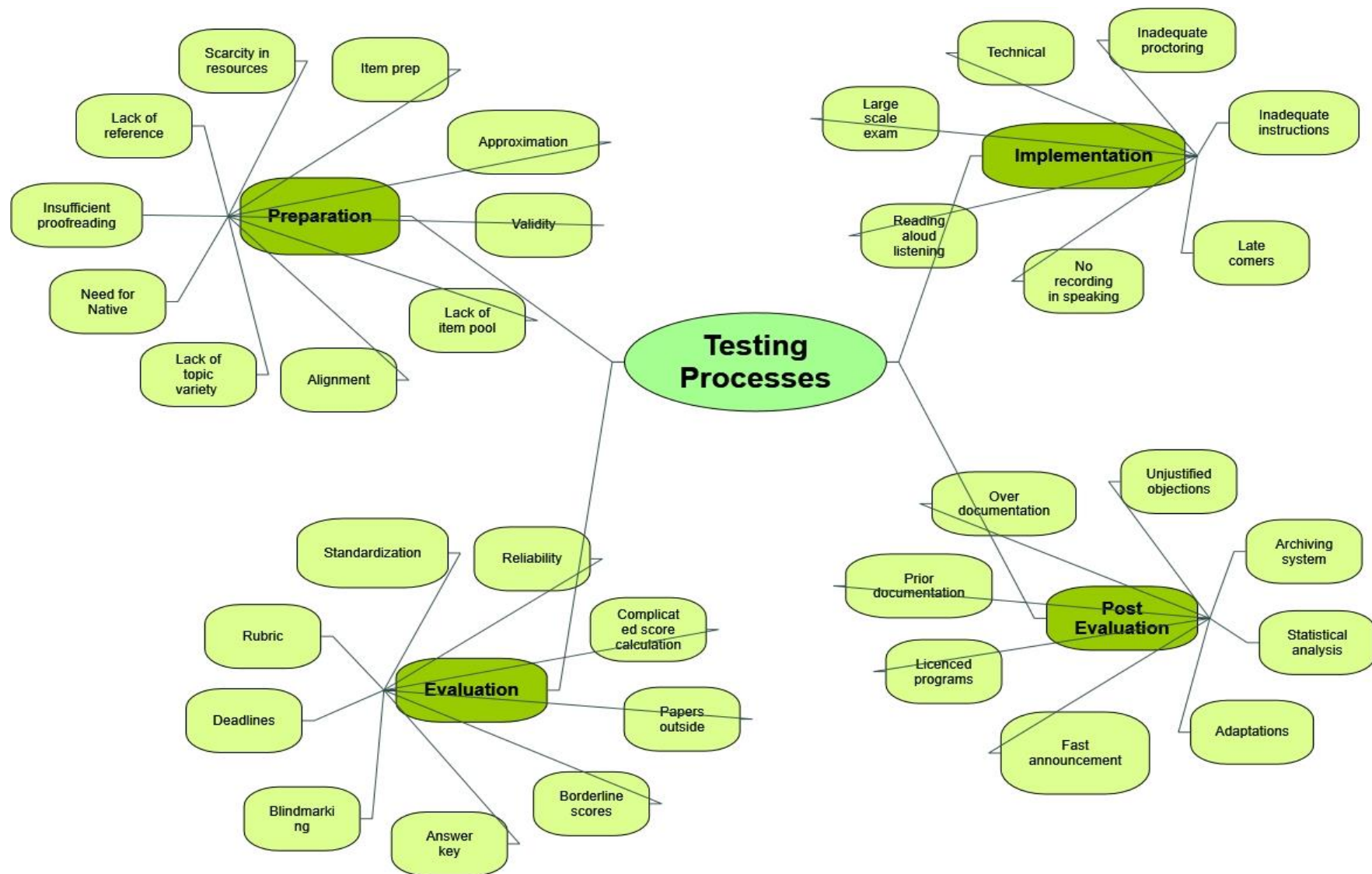


Figure 4.4. Testing processes-related Challenges

The next subsections deal with the challenges during preparation, implementation, evaluation and post-evaluation, explained respectively.

4.1.3.1. Challenges during preparation

This subcategory includes ten emerging codes which are shown in Table 4.12.

Table 4.12. Challenges during preparation

Emerging Codes	Number of Codes	%
Approximation to levels	131	13,86
Difficulty in item preparation	128	13,55
Scarcity in resources	124	13,12
Lack of reference framework	117	12,38
Insufficient proofreading	106	11,22
Need for native speaker	97	10,26
Validity matters	82	8,68
Alignment matters	73	7,73
Lack of item pool	52	5,50
Lack of topic variety	35	3,70
Total	945	100

The most frequent issue in this subcategory was the approximation of students' levels (N=131). One testing member mentioned in the interview:

Kişisel olarak language level yani öğrencilerin language levelını test etmek, language levelı tutturmakta çok zorlanıyorum. Tam bir sınav hazırladım diyorum, ancak ikinci okumada bu uygun mudur diye soru işaretleri beliriyor. Arada kalıyorum, içime sinmiyor...

Personally, I have a hard time preparing a test according to the language level of the students. Although I may think I have completed preparing the exam, in the second reading, question marks appear as to whether it is appropriate. I'm stuck and I don't feel comfortable.

The challenges related to difficulty in item preparation especially in listening (N=128) comprise the second most frequently used type in this subcategory in the interviews. This is one of the main challenges of testing members as indicated in the following extracts:

Listening bölümlerinde çok sıkıntı yaşıyoruz. Listening bulmak çok zor olabiliyor. Geçen de tartıştık. En sıkıntılı kısmı orası oluyor, listening sınavlarını native okuyor. Ama bu da mesela nativemiz okuduğumuzda tek bir aksan duymuş oluyor, başka bir aksan alternatifi

olmuyor. Dinledikleri kişinin pace yani hızı ile ilgili hocam. Elektronik ortamda mı? Bazen biz kendimiz okuyoruz. Hazır ready-made yok bizde. Ayrıca, eski kitapların CD'si yok ama bizim vizelerimizde listening bölümü var.

We're having a lot of trouble with the listening sections. Listening can be very difficult to find. We have recently had an argument. That is the most troublesome part, our native speaker reads the listening exams. But, when our native reads, students hear only one accent, there are no other alternative accents. Another side is about the pace, that is, the speed of the person they are listening to. Is it electronically recorded? Sometimes we read it out ourselves. We do not have ready-made exams. Also, old books do not have CDs, but there is a listening section in our midterms.

Another point was mentioned by a testing unit member:

Diğer bir sorun da authentic materyal oluşturmak ve çeldirici yazmak. Hazırladığımız sınavlar ne kadar dilin doğasına uygun ve authentic, emin değilim. Ayrıca, birçok sınavımız yani aslında bütün sınavlarımız multiple-choice olduğu için çeldirici yazmamız önemli, o yüzden çeldiriciler baya zamanımızı alıyor.

Another problem is creating authentic materials and writing distractors. I am not sure to what extent the exams we prepare are suitable for the nature of the language and authentic. Also, since we have many exams, in fact, all our exams in multiple-choice, it is important for us to write distractors, so they take a lot of time.

Scarcity in resources (N=124) has the third highest frequency among the challenges in this subcategory testing members face when codes are computed. A testing member elaborated on the issue as follows:

Kaynak sorunu, büyük bir çıkmaz bizim için. Doğru materyali bulma, artı doğru materyali modifiye edebilme çok zaman alıyor. Yeni kaynak alamıyoruz, kitaplar çok pahalı. Okula ayrılan bütçenin az olduğu, idare etmemiz gerektiği söylendi. İnternet açık kaynaklarını kullanmamızı müdürlük istemiyor, içerik güvenliğinden dolayı. Eleman sayımız az ve iş yükümüzün yoğun olduğundan yeni içerik üretemiyoruz. Burada güzel bir testing kütüphanemiz var ama geliştirme konusunda daha çok çalışmamız gerekiyor. Benimde kendi adıma yetersiz olduğum bölümlerden biri. Bunun üzerine gidip, araştırıp, daha iyi daha yeni kaynaklar sağlayabilir miyiz, sağlamalıyız diye düşünüyorum.

The resource problem is also a big dead-end for us. Finding the right material, plus being able to modify the right material, takes a lot of time. We can't buy new resources since books are very expensive. We were told that the budget allocated to the school was small and that we had to manage it. The directorate does not want us to use open sources on the internet considering the security of the content. We are unable to produce new content due to low staff and a heavy workload. We have a nice testing library here, but we need to work

on developing new materials. Indeed, I'm not good enough. I think we should do research and produce better and newer resources.

Another member explained further:

Kaynak problemimiz var bizim, kaynak bulmakta zorluk çekiyoruz. Çoğu zaman internete yöneliyoruz kendi çabalarımızla, o da her zaman problemlerimizi çözmeyebiliyor. Bu kitapların hepsi çok eski, senelerdir sınav komisyonunda olan kitaplar ve tükenmiş durumda artık içleri. Kendi çabamızla böyle yayınevlerinden bir iki kitap getirttik ama sürdürülebilir bir çözüm değil.

We have a resource challenge in finding resources. Most of the time, we tend to use internet sources with our own efforts, yet it does not always solve our problems. All of these books are very old, staying on the shelves of exam commission for years and are now fully used. We received one or two books from publishing houses, but it is not a sustainable solution.

As a number of participants indicated in the interviews, the lack of a reference framework leads to ambiguity during the preparation phase (N=117). One of the testing members explained her concerns as follows:

CEFR la genelde öğrenci seviyesi örtüşmeyebiliyor. Bu durumda, CEFR'a göre belirlenen dil düzeyi ile öğrencilerin düzeyi maalesef tutmadığı için hani öğrenciye odaklı mı yazmalıyız, yoksa o CEFR düzeyine mi odaklı yazmalıyız konusunda sıkıntılarımız var. Yani öğrenciye göre mi sınav hazırlayacağız, seviyeye göre mi sınav hazırlayacağız? En büyük sıkıntılardan biri. Bize sorarsanız, kendi içgüdülerimize ve çok dolaylı CEFR'a dayanarak diyebiliriz hocam. İçgüdüden kastım da iş tecrübesi ama hoca olarak. İllaki biraz simplificationa gidiyoruz. Eğer biraz idealist bir writersanız zaten içinize sinmiyor. Benim şahsen iç savaşım mesleğimle ilgili.

In general, the student level may not be in line with CEFR. In this case, since CEFR language level and the level of the students, unfortunately, do not match, we have problems with whether we should write tests focusing on the student or CEFR level. In other words, will we prepare the exam according to the student or the right CEFR level? One of the biggest challenges I have. If you ask us, we prepare tests based on our own instincts and indirectly on CEFR. I mean work experience, but as a teacher. We are doing a little simplification. If you're an idealistic writer, you may not be satisfied. In my profession, it is my personal civil war.

Another testing member added:

Seviyeler arası geçişler beceriler bakımından belli değil. Örneğin kelime listemiz bile yok, bizim bence en büyük sorunumuz farklı seviyelerde becerilerin gelişimini gösteren bir çerçevemiz yok.

Transitions between levels are not clear in terms of skills. For example, we don't even have a word list, I think our biggest problem is that we don't have a framework that shows the improvement of skills at different levels.

Another major concern of the directors and testing members related to national challenges is insufficient proofreading (N=106). One testing member expressed her concerns:

Yeterli ve düzgün işleyen bir proofreading döngümüz yok. Sınavı ben hazırlıyorum da nasıl hazırlıyorum, yapıcı dönüt olmayınca iyi mi yaptım, kötü mü tam anlayamıyorum. Önceki müdür zamanında biz aramızda bakıyorduk ama yeni müdür buna gerek olmadığını, sınavların geç kaldığını söyledi.

We don't have adequate and proper proofreading cycles. I am preparing the exam and I cannot fully understand how well or badly I am preparing it when there is no constructive feedback cycle. During our previous manager, we were checking each other's exams, but the new manager said that there was no need for this since the exams were late.

One of the sources of anxiety related to testing processes was testing members' stress about the need for a native speaker (N=97). As stated by a testing member:

Bir sürü hazırlık öğrencimiz ve sınavımız var. Tüm sınavlar için geçen yıl sadece bir tane proof-raderımız vardı. Çok yoğun dönemlerde yığılma olduğunda, proofreaderımız sürekli zaman konusunda problem yaşıyordu. Bu yıl böyle bir imkanımız bile yok. YÖK onaylamamış native hocamızın görevlendirmesini. Nasıl native proof'u yapacağız. Otorite kurum bizim sorunumuzun farkında bile değil.

We have a lot of students and exams in our school. We only had one proofreader last year for all exams. Our proofreader was constantly having problems with time when there was a piling of papers during peak periods. This year, we don't even have such an opportunity. The assignment of our native teacher was not approved by YÖK. How can we do native proof? The authority institution is not even aware of our problem.

Another challenge experienced by testing members was reported as validity matters (N=82). The following extract exemplifies how a testing member was affected:

Validity gerçekten vahim, "Are we asking what we are teaching?" bilmiyorum. Construct validity olarak, hem biraz sağlıyor hem sağlamıyoruz. Face validity bakmaya çalışıyoruz

ama çok yüzeysel. Content validity için öğrenciler “bize öğretilmeyen konular soruldu” diye şikayet ediyorlar. Ayrıca gerçekten bir öğrencinin ihtiyacı olan bütün skilleri, bütün becerileri onlara sağlıyor muyuz, hayır. CEFR’ın maddeleri ile “I can do statements”lar konusunda çok iyi olmadığımız açık.

Validity is really a disaster, “Are we asking what we are teaching?” I don't know. As construct validity, we provide some. We try to ensure face validity, but it's very superficial. For content validity, students complain about having been asked about topics that were not taught.” Also, we don't really provide all the skills, all abilities that students need. It is clear that we are not very good at employing CEFR and "I can do statements.”

As perceived by the participants, it seems that alignment matters seem to affect the quality of testing processes (N=73). A testing member was anxious about this challenge and she explained the situation in the interview as follows:

*Bazen seçilen kitapların, belirli seviyede o kitapları gören öğrencilerin ulaşmasını beklediğimiz seviyeye pek uyuşmadığını fark ettik. Ve bu durum testingi de dolaylı olarak etkiliyor tabii ki çünkü genel baktığımızda öğrencilerin seviyesinin altında kalıyor seçilmiş olan kitap. Hatta FOAI’de de konuşmuştuk kitap üzerinde B1 yazıyor ama B1 olduğunu kanıtlayıcı bir şeyleri yok. “Rapor isteyin” demişti uzman E*** Hoca hatta, bilmiyorum öyle bir raporları var mı? Dolayısıyla biz testing olarak dönem sonunda öğrencinin olması gereken seviye yani B2 - B2+ olmuş olması gerekiyor beklenti olarak. Ama kitap öğrenciyi o seviyeye getiremedi, kitabın kendisi o seviyede değil çünkü. Dolayısıyla biz alignment sağlamaya çalıştığımızda derste işledikleri içerikle sıkıntı yaşıyoruz o seviyeyi yakalamada.*

We have noticed that sometimes the selected books do not quite match the level we expect students to reach who have seen those books at a certain level. And this situation indirectly affects testing, of course, because generally speaking, the chosen book is below the level of the students. We even mentioned this at FOAI, B1 is aimed at the book, but they have nothing to prove that it is B1. “Ask for a report,” the expert E*** suggested, yet I don't know if they have such a report. Therefore, at the end of the term, we expect the student to be at the required level, namely B2 - B2+. But the book could not bring the student to that level, because the book itself is not at that level. Therefore, when we try to provide alignment, we have trouble with the content they teach in the course, and reach that level.

A testing member agreed and added some more details:

Table of specification her yerde çok da sağlıklı hazırlanmıyordur diye düşünüyorum. Bir test planının yapılma süreci önemli olmasına karşın genelde bizim toplum planlamayı çok sevmiyor, direk hadi Allah deyip başlıyor. Table of spec geçen toplantıda paylaştım ama çok sıcak bakmadılar. Yani biz zaten biliyoruz, ne gereği var gibi baya açık bir şekilde böyle söylemişlerdi. Alaylı olduğumuz için kafamızda var, ama yazılı değil.

I think the table of specifications is not prepared appropriately everywhere. Although the process of preparing a test plan is important, our society does not like planning very much, they just start by citing the name of "God." At the last meeting, I shared the table of specs, but they were not interested. I mean, they claimed they already know it quite clearly and questioned if it was necessary to murmur all the plan is in their heads because they graduated from a university. They believe they have it but it's not written.

Some of the testing members indicated that anxiety appeared due to a lack of the item pool; in other words, item bank (N=52). A testing member, for instance, reported:

Kurumların soru bankaları genelde yok, var diyenlerinki de etkin ve verimli çalışmıyor. İstenilen düzeyde değil bence.

Schools generally do not have item banks, and those, claiming they have, do not have effectively and efficiently working banks. I don't think it's at the desired level.

A testing member added some more details and elaborated on the source of the challenge as follows:

Soru havuzu için tabi biraz istatistik çalışmak gerekiyor, biz ona başladık ama bir software e ihtiyaç duyduk. Okulumuz onu birazcık pahalı buldu, almadı. E almayınca biz de kendi çabalarımızla bir şeyler yapmaya çalıştık hani. Kullanılan sorunun nerede kullanıldığını görmemiz lazım, onun kaydını nasıl tutacağız bu yoğunluğun arasında, olmadı o yani. Aslında çok istedik ve çalıştık onun üzerine. Hakikaten böyle her annual toplantı sonrası item banking hep bir maddeydi bizim için. E tam başlıyoruz software ihtiyacı çıkıyor karşımıza. Olmadı falan sonra bıraktık, dedik herhalde bu böyle olmayacak artık yani. İdeali o tabii ki biliyoruz. Biz soruları yazıyoruz, hani iyi çalışan sorular ama öyle kalıyor. Diğer bir nokta ise belki aralarında kötü çalışan sorular var ve kullanılıyor. Elenmesi gereken. O anlamda eksikiz.

Of course, some statistical work is required for item pool. We started to work on it, but we needed software. Our school found it a little expensive and did not buy it. When we didn't get it, then we tried to do something with our own efforts. Due to the fact that we needed to see where the question was used, how we were going to keep track of it, etc., we could not realise it when we had such a busy schedule. Actually, we wanted a lot and we worked on it. Indeed, at every annual meeting, item banking has always been a matter of discussion. Well, when we start to discuss and the need for software appears. In the end, we quit admitting the impossibility of item banking. We are aware that it is the ideal. We write items, you know, which work well, but we lose them. Another point is that maybe there are some items that work poorly and they are re-used though they must be eliminated. We are insufficient in this issue.

One final reason seems to be a lack of topic variety (N=35). One testing member explained how she felt:

Öğrencilerin düşük seviyelerine bağlı olarak farklı konuları sınıflara taşıyamıyoruz. Bu sorun aslında K12'den gelen bir sorun. Çocuklar sınıfa getirdiğimiz güncel olmasına karşın bu konuda destekleyici fikir çıkartamıyorum diyor. Major idea, minor ideaları oluşturmuyorlar. Sosyal medya hesaplarının da etkisi olabilir, bir iki cümle söyleyip konuyu bitirmek istiyorlar.

Due to students' low level, we are unable to bring different subjects into the classrooms. This issue actually results from K12. Students complain they can't come up with supporting ideas even though we bring up-to-date topics to the classroom. They can't produce major ideas or minor ideas. They want to end their turn by saying one or two sentences, which may be caused by social media use.

Challenges during preparation were related to the lack of training since the TUMs who were forced to work in the TUs stated that they had no or limited knowledge of how to design a language test. The most frequently mentioned challenge was approximation of language tests to students' levels. It was claimed that they had difficulty preparing items and determining item difficulty and distractors. Besides, scarcity of resources, lack of reference framework, insufficient proofreading cycle, lack of a native speaker, and validity and alignment matters were other major challenges preventing them from preparing good language tests. The next subsection deals with challenges during implementation.

4.1.3.2. Challenges during implementation

This subcategory includes seven emerging codes as shown in Table 4.13.

Table 4.13. Challenges during implementation

Emerging Codes	Number of Codes	%
Large scale exams	153	21,89
Technical issues	145	20,74
Inadequate proctoring	139	19,89
Inadequate instructions	102	14,59
Latecomers to exams	76	10,87
No recording in speaking exams	51	7,30
Reading aloud listening parts	33	4,72
Total	699	100

Large scale exam (N=153) has the highest frequency among the challenges in this subcategory testing members face when codes are computed. A testing member explored as follows:

Eş zamanlı çok fazla sınıfta sınav yapıyoruz. Genelde eş zamanlı 30 sınıfta falan olabiliyor. Bazı sınıflara ek süre veriliyor, acil çözüm bulmanız gereken durumlar oluyor. Genelde bu tip problemler.

We conduct exams in too many classes at the same time. In general, it can be done simultaneously in 30 classes. We need to find urgent solutions for unexpected situations. For instance, some classes are given additional time. We have such challenges.

Another testing member added:

Sınıf sayısının yanı sıra öğrenci sayısının çok fazla olması da bence büyük sıkıntı. Sınıfta çok öğrenci olması sekreteryaya işini çok fazla artırıyor. Yani biz her seferinde çok büyük ölçekli sınavlar yapıyoruz aslında. Mesela 3000 küsur öğrenci var. Özellikle en yoğun eylül ayıdır. Eylülide en kapsamlı muafiyeti yapıyoruz. Dönem başında 3500 falandı. Yıl sonuna doğru 2500-2600 civarında öğrenci sınava giriyor. Sınıfların kalabalık olmasından dolayı grup oluşturmak zorunda kalıyoruz.

In addition to the high number of classes, the large number of students within the classroom is also a big challenge. Having many students in the class increases the secretarial work too much. In other words, we always conduct a number of large-scale exams. For example, there are more than 3000 students. Especially the busiest month is September. We make the most crowded exemption in September. At the beginning of the term, it was around 3500. Towards the end of the year, it decreases to around 2500-2600 students. We have to form groups because the classes are crowded.

The second most frequent anxiety in this subcategory was testing members' concerns about technical issues (N=145) such as malfunction of photocopy machines, speakers, audio systems, and USBs. One testing member stated in the interview:

Teknik problemler yaşıyabiliyoruz. Teknik problemler derken, listening ile ilgili, smartboard ile ilgili problemler olabiliyor, ses kalitesi düşebiliyor. Öğrencileri zamanında sınıflara getiremediğimiz durumlar olabiliyor. Çok nadir de olsa hava durumuyla ilgili problemler olabiliyor, geç başlatma durumunda kalabiliyoruz. Uygulama esnasında proctorlarımızla, görevli arkadaşlarımızla ilgili zaman zaman problemler çıkabiliyor. En basiti listening bölümünde teknoloji biraz devreye girince teknik arızlar olabiliyor. O durumda geç başlıyor listening, biraz ek süre veriliyor. Şu anda aklıma gelenler bunlar.

We may experience technical problems. There may be problems with listening, the smartboard, and the sound quality may decrease. Sometimes we cannot bring students to

class on time. Although rarely, there may be problems with the weather. Then we may have to start late. During the application, problems may arise from time to time with our proctors and our colleagues. The simplest is listening as technology comes into play when technical malfunctions may take place. In that case, listening starts late, and extra time is given. These are the ones that come to my mind right now.

Two testing members from another university explained further:

Projektörler eski çalışmıyor. Teknoloji sonuçta çabuk yıpranan bir şey ve çabuk bozulan bir şey. Sınıfımızda tabii ki hoparlörümüz var ama bir hafta sonra çalışmayabiliyor. Yani sürekli yenilenmesi gerekiyor ve tabii ki bunlara da para harcanması gerekiyor. Bütçemiz ise çok sınırlı, bakan arkanızdaki fotokopi makinesi bozuk, çalışmıyor. Şu anda final sınavını basamıyoruz mesela. Ayrıca buradaki optik okuyucu bozuk, optik kağıtları okuyamıyoruz. Yani fiziksel koşullarımız çok zor. Bu makineyle 8000 öğrenciye on sayfalık bir sınav hazırlamak aslında olmaz. Senin ortasından beri biz bu makineyle çok büyük sorun yaşıyoruz. Bastıklarının yarısı okunmuyor, baştakiler değil arada okunmayanlar var. Bu yüzden baştakileri de tek tek çekiyoruz. Hangisi okunuyor, hangisi okunmuyor anlamak bile çok zaman alıyor. Aradan hiç beklemediğimiz bir kağıttan çıkıyor. Dediğim gibi teknoloji çabuk yıpranan bir şey olduğu için, çok eskidiği için, sürekli bozulma sayısı da artıyor.

The projectors are old and not working. Technology is, after all, something that wears out quickly and is a perishable thing. We have a speaker in our classroom, but it may not work after a week. In other words, they need to be constantly renewed. Money needs to be spent on them. However, our budget is very limited. Look, that copy machine behind you is broken, it doesn't work. For example, we can't print the final exam at the moment. Also, the optical reader here is broken, so we cannot read optical papers. Our physical conditions are very difficult. With the copy machine, it is not possible to prepare a 10-page exam for 8000 students. Since the middle of the year, we have been having this problem with the machines. Half of the prints are unreadable. The unreadable ones are not the first ones. That's why we need to copy them one by one. It takes a lot of time to even understand which one cannot be read. We cannot expect which paper. As I said, since technology is something that wears out quickly and gets old, the number of breakdowns is constantly increasing.

*Makinelerimiz çalışmamasına ek olarak, teknik işlerle, teknik meselelerle ilgilenen bir görevli yok aslında. Yani bir-iki görevli var ama onların da görev tanımında öyle bir şey olduğunu sanmıyorum. Ve bu konuda da çokta iyi değiller, öyle hemen çözüme ulaşacak bilgileri, eğitimleri yok. Birimizde çalışan M***** Bey'den bir şey istediğimizde tabii ki de gelip çözmeye çalışıyor, ama o da bilgisi dahilinde bize yardımcı olabiliyor. Daha*

fazlasını yapamıyor. Servis çağırılıyor, ama servisin gelmesi uzun zaman alıyor. Birimde IT alanında uzman, böyle teknik bir kişinin de görevlendirilmesi gerekiyor açıkçası burada.

In addition to the malfunctioning of our machines, we lack a staff who takes care of technical repairs and issues. There are one or two members, but I don't think there is such a thing in their job description. And neither are they good at technical matters, nor do they have the knowledge or training to reach a solution immediately. When we want something from Mr. M*****, who works in our unit, of course, he comes and tries to solve it, but he can help us with his knowledge in a limited way. He cannot do more. The service is being called, but it takes a long time for the service to arrive. Obviously, such a technical person, who is an IT expert, should be working here.

A testing member shared similar views but also put emphasis on the desperate situation of the school:

Bir üniversitede YDYO vitrindir aynı zamanda. Ancak öğrencilerin üniversiteye ilk adım attığında eğitim almak üzere geldikleri okullarda şartlar çok zayıf. Mesela “fotokopi makinemiz arızalandı” diye İİBF’ye gittim, “fotokopi çekebilir miyiz” dedim. Ama izin vermediler. Fotokopi çekeceğim öğrenciler aslında o fakültenin kendi öğrencileri. Yani bir sene sonra hazırlığı geçerek o fakülteye gidecek öğrencilerin kağıtlarını çekemedim. Dışarda kendi paramla çektirdim.

YDYO is a display window in a university. However, the conditions at schools where students come to receive education when they first step into university are very poor. For example, I went to the faculty of economics and administrative sciences because "our photocopy machine broke down." I asked, "Can we copy exam papers?". But they did not allow it. The students I will copy exam papers for are actually the students of that faculty. In other words, I could not copy the papers of the students who would pass the preparation and go to that faculty in a year. I had to do it by spending my own money.

The challenges related to inadequate proctoring, including missing packs, late handing of packs, tendency to take medical reports and day-off (N=139) comprise the third most frequently used type in this subcategory in the interviews. A testing member, for instance, considered:

Sınav esnasında ve sınav sonrası farklı uygulamalar yapıldığını duyuyoruz ya da şahit oluyoruz zaman zaman. Ve buna yapabileceğimiz bir şey olmuyor. Sınav esnası ile ilgili, mesela biz instruction’ımıza, prosedürümüze şunu koyduk; kesinlikle Türkçe konuşulmayacak, sorular açıklanmayacak. Ama hoca kapısını kapatıp bunu yapabiliyor yani. Proctorlar optional, yoruma açık sorularda eşit davranmayabiliyorlardı. Öğrencilerine acıyıp eğer anlamadıysa açıklıyorlar falan. Soruyu açıklamakla yetinmeyip

örneklendirebiliyor mesela. Sınav sonrasında ise kağıtları sıralamayanlar veya geç getirenler oluyor. Geçen dönem kağıtları eve götürmüş hocamızın biri, bir kağıt bulunmadı.

From time to time, we hear or witness that different applications are made during and after the exam. And there is nothing we can do about it. For example, we put the following in our instructions, our procedure; no Turkish will be spoken, and questions will not be explained. But some instructors can close the door and do so. Proctors could behave equally in optional, open-ended questions. They take pity on their students and explain if they don't understand. For example, some not only explain the question but also give an example. Some instructors do not sort the papers or bring them late. One of our instructors took the papers home the last term, and could not find one paper.

A testing member added some more details and elaborated on the issue as follows:

Hocalarımız sınavda gözetmenlik yapmak istemiyorlar, izin istiyorlar, vermeyince de rapor alıyorlar. Bu durumda diğer fakültelerden asistan falan geliyor gözetmenlik için. Hamile olan hocamıza kolaylık yapmak istedik, başka bir hocamız geçmişte hamile iken gözetmenlik yaptığını ama şimdi çocuklarının büyüdüğünü belirterek ayrımcılık yapıldığını iddia etti. Başka bir hoca da erkek olduğu için hamile kalamadığını ifade etti. Şaka gibi diyaloglar yaşadık.

As our instructors do not want to take duty in the exams, they ask for permission. When we do not let them, they take a medical report. In this case, assistants from other faculties are assigned. We wanted to help a pregnant instructor, another instructor claimed that she had examination duties in the past when she was pregnant, but now her children are grown, claiming this as discrimination. Another instructor stated that he could not get pregnant because he was a man. We had such conversations like jokes.

An important number of interviewees indicated that the source of the challenges encountered in testing processes results from inadequate instructions/guidelines (N=102). The following extract explains her feelings:

Bazen sorun instructionlarla ilgili olabiliyor. Sınavın yönergesindeki sorunlar özellikle yazma için diyebilirim. Yani yazma yönergeleri bazen sınavda pek yeterli olmayabiliyor. Sınav esnasında bunu açıklamak şansınız da olmuyor.

Sometimes the problem can be with the instructions. The instructions of the exam especially in writing can be problematic. In other words, writing instructions are sometimes not adequate for the exam. We do not have the chance to make explanations during the exam.

Furthermore, latecomers to exams (N=76) are a noticeable concern of testing members. One testing member expressed her feelings:

Çok fazla geç kalan öğrenci oluyor. Geç gelenleri almayalım diye karar aldık ama veli şikayetleri ve rektörlük kararımızın kaldırılmasını istedi. Geç kalan öğrenciler, diğer öğrencilerin konsantresini fazlasıyla etkilemekte.

Too many students are late. We decided not to accept latecomers, but upon parents' complaints, the rectorate demanded that our decision be abolished. The students who are late greatly affect the concentration of other students.

No recording in speaking exams (N=51) also affected the motivation of testing members in a negative way. A testing member, for instance, explained the issue as follows:

Speaking'leri kaydedemediğimiz için öğrenciler itiraz edemiyor.

Students cannot make objections because we cannot record the speaking part.

Finally, some testing members were anxious when they had to read aloud listening parts (N=33). A testing member elaborated:

Listening hocalar okuyor, speakerlarımız bile yok yani. Bu devirde olmaması çok enteresan ama işte olmayınca da o sorunları aşmak, kişilere kalıyor. O zaman da kişiler burn out dediğimiz şeye ister istemez yakalanıyor.

Instructors read aloud the listening parts since we don't even have speakers. It is very interesting not to have speakers in this era, but when there are not, it is up to people to overcome those problems. Then, people inevitably get caught in what we call burnout.

The results revealed that the highest-ranking challenge in this category was large-scale exams, resulting in heavy workload ranging from photocopy duties to delivering and receiving packs from the instructors. Another major challenge was about technical challenges, including malfunction of machines such as photocopy, speakers, audio system, USBs, etc. The other most striking problems included inadequate proctoring, insufficient instructions and guidelines. These problems are followed by latecomers to exams, no recording in speaking exams, and having to read aloud in listening exams. The next subsection deals with challenges during evaluation.

4.1.3.3. Challenges during evaluation

This subcategory includes nine emerging codes which are shown in Table 4.14.

Table 4.14. Challenges during evaluation

Emerging Codes	Number of Codes	%
Reliability issues	224	19,48
Difficulty in standardisation	187	16,26
Lack of well-developed rubric	161	14,00
Failing to meet assessing deadlines	143	12,43
No blind marking	135	11,74
Unclear answer key	114	9,91
Borderline scores	92	8,00
Assessment of papers outside the school	63	5,48
Complicated score calculation	31	2,70
Total	1.150	100

The challenges related to reliability issues such as single grader/assessor in writing and speaking (N=224) comprise the most frequently used type in this subcategory in the interviews. A testing member elaborated on the issue and stated that:

Reliability sorunu sadece bizim üniversitede değil, neredeyse belki bazı istisna üniversiteler hariç tüm devlet ve vakıf üniversitelerinde büyük bir sorun. İnsanlar mevcut iş yüklerine ek hiçbir ekstra iş istemiyorlar. Zoraki verdiğimizde de çok üstünkörü bir şekilde yapıyorlar. Writing kağıtlarını iki kişiye okutamıyoruz, sürekli şikayet ediyorlar. Speaking de sınırlı personeli match etmek oldukça zor. Kaliteyi artırmak istiyoruz, ancak mevcut kapasite ve iş yüklerine baktığımızda bir şey diyemiyoruz. Bu gibi görevler ekstra algılandığından pek de umursanmıyor.

Reliability issue is a big problem not only in our university but also in almost all state and foundation universities with the exception of some. People don't want any extra work in addition to their current workload. When we force them to get it, they do it in a superficial way. We cannot have two people to score writing papers, they are constantly complaining. It is quite difficult to match the limited staff for the speaking part. We want to enhance the quality, but considering the current capacity and workloads, we can't do anything. As such tasks are perceived as extra, they are not paid attention much.

Difficulty in standardisation (N=187) has the second highest frequency among the challenges in this subcategory testing members face when codes are computed. Some testing members explained as follows:

Değerlendirmede uyumsuzluklar yaşanabiliyor. Atıyorum nerede, ne kadar puan kırılmalı-kırılmamalı o noktada bazen eksik kalabiliyoruz. Fikir ayrılıkları oluyor çünkü herkes farklı düşünüyor, o fikir ayrılıklarını yönetebilmek de sıkıntı. Özellikle writing değerlendirmesinde farklılıklar olabiliyor, her hocanın farklı bir tabii ki doğal olarak kişisel farklılıklardan dolayı standartlaşma her zaman mümkün olmayabiliyor.

There may be discrepancies in the assessment. We lack training on where and how many points we should take some points off. There are differences of opinion because everyone thinks differently. Managing those differences is a problem. There may be differences, especially in the evaluation of writing as each rater is different. Of course, naturally, due to personal differences, standardisation may not always be possible.

Değerlendirmedeki endişelerden dolayı sınavlarda açık uçlu sorularla ilgili çok sıkıntı yaşıyoruz. Standardı sağlamak çok zor oluyor. Bu durumda sınav türleri arasında avantaj dezavantaj çıkmış oluyor. Bu kadar fazla öğrenci ve bu kadar az kişi ile açık uçlu soruları ve writing kısmını istediğimiz gibi yapamayacağımız için biz de açık uçlu sorulardan kaçınıyoruz. Yani çok fazla alternatif çıkacağını düşünerek, bilinçli olarak writing ve açık uçlu soruları sınavlarımızdan çıkardık ve multiple-choice'a yöneldik. Aslında çok mantıklı değil ama öyle bir kısıtlamaya gittik.

We have a lot of challenges with open-ended questions in exams due to concerns in assessment. It is very difficult to maintain standardisation. In this case, there are advantages and disadvantages among exam types. Since we cannot do the open-ended questions and the writing part as we would like considering so many students and so few people, we avoid open-ended parts. In other words, assuming that there would be too many alternatives, we deliberately removed the writing part and open-ended questions from our exams and tended to employ multiple-choice. Actually, though it doesn't make much sense, we took such a restriction decision.

Değerlendirme aşamasında tabi hocaların bakış açısına göre, dil eğitimine göre kağıtları yorumlamaları da farklı oluyor. Tamam hepimiz iletişimsel metot biliyoruz ama iç sesimiz farklı olabiliyor. Her zaman aynı olmuyor, bırakın açık uçlu kısımları mesele dinleme ile ilgili bir metni değerlendirirken dahi kimi hoca diyor ki işte bir harfi yazmamış, sıfır not verelim. Aslında ben şeyden bakıyorum perspective of communicative testing yani olanı önce bir görelim. Bireysel farklılıklardan dolayı bu tip bazen değerlendirmedeki method yaklaşımları gibi sorunlar olabiliyor. Hocalar arasında o yüksek vermiş, o düşük vermiş tartışmaları olabiliyor.

During the evaluation phase, of course, the interpretation of the papers differs according to instructors' points of views and language education level. Well, we all know though the communicative method, our inner voice could be different. It is not always the same. Let alone the open-ended parts, even when evaluating a text about listening, some instructors say when students don't write even a letter, they want to give a zero grade. Actually, from my perspective, what is important is communicative testing, that is to see what happens first. Due to individual differences, evaluation can sometimes cause problems such as differing scoring approaches. There can be arguments among the instructors on some give high or others give low.

Some testing members shared similar views but also put emphasis on the responsibility of instructors:

Hocalar kağıtlarına göz atıp gelmeleri gerekiyor toplantıya, ancak bazen öyle gelmiyorlar bazı sınavlardan sonra toplantı yapıyoruz hani istiyoruz ki hocalar kağıtları gözden geçirsin de sorunlu kağıtları konuşalım diye. Hiç bakmayan hocalar, toplantıya gelmeyen olabiliyor. Ama sonradan da sıkıntı çıkabiliyor. Karşılıksız çek gibi görüyorlar ve görev tanımında olmadığını düşünebiliyorlar.

The instructors are expected to come to the meeting by having a look at the papers, but sometimes they don't do so. We organize meetings after some exams. You know, we want the instructors to review the papers beforehand so that we can solely talk about the problematic papers. There may be instructors who have never had a look or do not even come to the meeting. But later on, challenges may arise. They may consider it as an unpaid check and that it is not defined in their job description.

Rubric oluşturduk ama zaman zaman arkadaşlarımız refer etmiyorlar ona, gözden kaçırmış olabiliyorlar. Çok alıştık zaten hep aynı şeyler deyip tekrar tekrar ellerine almayabiliyorlar, kaçınıyorlar bundan. Onun dışında bazen rubrici doğru yorumlayamama olabiliyor, özellikle de yeni gelen arkadaşlarda. Oryantasyon yapamıyoruz, zaman zaman dönem ortası girişlerde oryantasyona vaktimiz kalmadığı için kaçamaklar olabiliyor. O yüzden de rubric tanıtılmadığı için hatalar oluyor.

We have formed a rubric, but from time to time, our friends do not refer to it, they may overlook it. They may think that they are so used to it and don't want to take it in their hands again and again, so they avoid it. Apart from that, sometimes some may have difficulty interpreting the rubric correctly, especially new members. We are unable to do orientation when new instructors start to work at mid-term. That is why we do not have enough time for orientation. Then there are misuses of rubric because it is not introduced.

A testing member added that there was a lack of coordination between graders and standardisation meetings were of great importance and explained:

Eskiden standardizasyon toplantılarımız vardı. İşte bu fiziksel koşullar altında onu da etkiledi. Her sınav sonrasında o bölümün o levelın hocalarıyla bir revise key toplantıları yapılıyordu ve alternatif cevaplara birlikte karar veriliyordu hocalarla birlikte. Toplantıların sıklığı ihtiyaca göre değişiyordu. Standart bir zaman yok. Ama şimdi iki ayrı kampüste olduğumuz için o iki grubu bir araya getirmek imkansız hale geldi. Bir de sayılar gitgide arttı tabii. Yeni hocalara standardizasyon toplantısı yapamıyoruz. Aslında çok büyük bir değişiklik yapıldığı zaman mesela kriter değiştiği zaman, bütün hocalara toplantı düzenliyoruz ancak gelmiyorlar o ayrı. Zaten bu kadar hocayı bir araya getir, yap et falan kalabalık nüfusla zor. Bu yüzden standardizasyon anlamında sürekli sıkıntılarımız oluyor. Yani tabi gönül istiyor her sınav öncesi toplantı yapalım, tartışalım, konuşalım ama tabi öyle bir vaktimiz olmuyor çünkü öncesinde hani hem inputu hazırlamanız gerekiyor, bizim bir bakmamız gerekiyor, sonra dönüt vermemiz gerekiyor falan. Bu yani, genel olarak hani en baskın aa yine mi olmadı dediğimiz bu.

We used to have standardisation meetings. The physical conditions actually affected it as well. After each exam, a revision key meeting was held with the instructors of that level and alternative answers were decided together. The frequency of the meetings varied according to needs. There was no standard time. But now that we are on two separate campuses, it has become impossible to bring those two groups together. And, of course, the number of instructors increased. We cannot hold a standardisation meeting for new ones. In fact, when a big change is made, for example, when the criteria change, we need to organize meetings for all the instructors, but they do not come. It's hard to get so many instructors together anyway, as a large population. That's why we have constant problems in terms of standardisation. Of course, we desire to have a meeting, discuss and talk before each exam, but we don't have such time because we have to prepare the input first, check it, and then give feedback. This is, in general, the most dominant part that hasn't been established yet.

The third most frequent anxiety in this subcategory was testing members' concerns about a lack of a well-developed rubric (N=161). A testing member elaborated:

Önceden analitik rubric kullanıyorduk ama sonradan holistic bakmanın daha rahat olacağı, daha doğru olacağı düşünüldü. Buradaki asıl sorun birçok hocanın kafasında var bir rubric. Şahsen ben kendi adıma ben dahi okurken kafamdaki rubric'e göre okuyorum. Geliştirilen analitik rubric eski hocalar ve yönetim tarafından çok iyi işe yaramadığı düşünüldü. Analitik rubric'in zayıf kaldığı ve öğrencilerin çok yüksek notlar aldığı düşünüldü. Özellikle B2 düzeyine yaklaşılmaya başladıkça rubric'in gözden geçirilmesi

gerekiyordu. Yani kağıt okudukça “bu kağıdın bu notu almaması gerekir” diyorduk yani. Sanırım o sebep oldu biraz da rubric’in kalkmasına ama hala ben B2 düzeyinde kağıt okuduğumda nasıl yani yanlış oldu herhalde, hala yüksek not verdiğimizizi düşünüyorum writing kağıtlarına. 20 üzerinden 17-18 alan kağıtlar aslında 17-18’lik değil, B2 seviyesi düşünüldüğünde. Proficiency için kullandığımız rubric aslında tepeden indi öyle olabilir.

We used analytical rubrics before, but later it was thought that it would be more beneficial and accurate to employ holistic one. The main problem here is the rubric that many teachers have in their minds. Personally, I score the papers according to the rubric in my head. The analytical rubric that was developed was considered by experienced instructors and management to not work very well. It was thought that the analytical rubric was weak and the students got very high grades. The rubric needed to be revised, especially at B2 level. In other words, as we scored the paper, we were saying that "the paper should not receive this grade." I think that caused the rubric to be abolished, but I still think it was wrong when I score a B2 level paper. I still think we gave high marks on some papers. The papers that got 17-18 out of 20 were not actually 17-18, considering the B2 level. The rubric we use for proficiency actually was accepted as top-down.

Among the challenges, instructors’ failing to meet assessing deadlines (N=143) was a great concern for testing members. A testing member, for instance, put an emphasis on deadlines and stated:

Deadline’lar falan yok. Final sınavı için mesela belirlediğimiz deadline’a bile uyulmadı. Belli bir süre veriyoruz, okuyup geri getirmeleri için ama kimse uymuyor. Ufak tefek geç kalmalar genelde aşıyor, ama mesela bir sene dolabında tutabiliyor sınav zarfını ondan sonra getirebiliyor. Kontrol edemiyoruz yani çünkü o kadar çok sınav evrakı var ki onu kontrol edemiyoruz. İmza karşılığı alıyoruz aslında ama bazen gözden kaçabiliyor. Yani dediğim gibi bir sınavı bitiriyorsunuz, diğeri başlıyor. Sürekli devam ediyor o yüzden kontrol edemiyoruz burada. Hocanın bir hafta süresi oluyor, hafta boyunca. Sınav evrakları evine gidiyor yani. Bazı sınavlar çoktan seçmeli ve optik okuma olduğu için biz burada okuyoruz. Deadline’a uymak da istemiyorlar, zor oluyor onlar için bazen. Sürekli deadline’ı geçiriyorlar, deadline’ı çok aksatıyorlar. Ayrıca burası sürekli bir sirkülasyon halinde, bizim buradan sınavı toplayıp arşive indirmemiz lazım ki yeni sınavı yerleştirelim. Ama hala getirmeyen hoca oluyor sınavını, biz arşive gönderemiyoruz sınavları. Sürekli hocayı bekliyoruz getirsin diye. Başkasına bağımlı çalışmak zorunda kalıyorsun.

There are no deadlines. For example, the deadline we set for the final exam was not complied with. We give a certain amount of time for the instructors to score papers and bring them back, but no one does so. Minor delays are usually compensated, but for example, one instructor kept papers in his locker for a year and then brought them afterward. We cannot control it because there are so many exam documents. We actually hand them through signature, but sometimes it can be overlooked. So, as I said, as we finish

one exam, another begins. It goes on all the time, so we can't control it here. The instructors have a week. Yet the exam papers are taken home. The exams that are multiple-choice or scored in the optical reader are scored in the school. They don't want to follow the deadline either, it's hard for them. They are constantly skipping deadlines. Also, this office is in constant circulation, we need to collect the exams and send them to the archive so that we can place the new ones. Since the instructors do not bring the exam papers, we cannot send them to the archive. We are constantly waiting for the instructors to bring them. We have to work dependent on someone else.

A number of participants indicated in the interviews no blind marking (N=135). A testing member cited her concern:

Writing ve speakinglerde kör hakem uygulaması yapalım diyoruz ama istediğimiz gibi olmuyor. Genelde değerlendirciler puanlama yaparken writing de yan yana okuyorlar, speaking de ise aynı puanı veriyorlar. Arada hiç farkın olmaması güzel ama sistemi bu niyetle kurmadık.

We try to set blind marking in writing and speaking, but it doesn't work as we want. Generally, assessors score together in writing, and they give the same score in speaking. It is fine that there is no difference, but it was not our intention when we set up the system.

As one of the nationwide problems of language testing, unclear answer key (N=114) caused anxiety among testing members in this study. The following extract from testing members' experience would be enough to clarify the issue:

Cevap anahtarları düzgün hazırlanmıyor, aceleye geliyor. Optik okuyucunun cevap anahtarında bir hata, her sayfada standart hata çıkarıyor. Açık uçlu sorularda ise sürekli değişen bir key var. Toplantı yapılmadığından fikir birliği sağlanamıyor, sadece aklına gelen WhatsApp'tan yazıyor. Düzenli bir sistem oluşturamadık.

Answer keys are not prepared properly but in a rush. An error in the answer key of the optical readers produces standard errors on each page. Keys for open-ended questions, on the other hand, constantly change. Since there is no meeting, no consensus can be reached. Instructors write what comes to their minds on WhatsApp. We have not created a regular system.

In addition to the concerns stated above, borderline score (N=92) is a factor contributing to anxiety. As explained by the participants:

Bizde sınavın yapılması, uygulanmasından ziyade sonrasında öğrenci itirazları ekstra iş yükü getiriyor. Çok fazla öğrenci itirazı alıyoruz. Sınırdaki kalan öğrencilerimiz için kurtarma operasyonları yapmak zorunda kalıyoruz. Bizim 65 geçme notumuz var ama 60-

64 aralığında alan, zor durumda olan öğrenciler için ek bir sınav hakkı veriyoruz zaten. Sene başından beri hazırladığımız o kadar sınav yetmiyor, bir sınav daha hazırlıyoruz. Neyse ki herkese ikinci şans mümkün değil ama bazı öğrenciler için bile olsa yeni bir sınav hazırlaması işte.

In our school, student objections bring extra workload, rather than taking and applying the exam. We get a lot of student objections. We have to carry out rescue operations for our students at the borderline. Our passing grade is 65, but we give an additional exam right for the students who are in the 60-64 range and who are in a difficult situation. In addition to many exams we have prepared since the beginning of the year, we are to prepare another exam. Fortunately, not everyone has a second chance, but even for some students, we are preparing a new exam.

Sınav sonuçlarının duyurulması sırasında bir problem yaşamıyoruz ama sonrasında itirazlar yağmur gibi geliyor. Öğrenciler için geçme notu 65 olmasına karşın 60-64 alan öğrencileri sınırda kalmış sayıyoruz. Bu durumda onlara tekrar bir sınav hakkı veriyoruz ikinci bir şans tanıyarak. E bunlar da bize hep yeni sınav olarak dönüyor. İşin özü 60-65 arasındaki borderline öğrenciler için tekrar bir sınav veriyoruz.

We do not have a problem during the announcement of the exam results, but we receive objections a lot. Although the passing grade for students is 65, we consider the students who get 60-64 to be at the limit. In this case, we give them the right to take another exam, giving them a second chance. And in the end for us, it is a new exam. In brief, we give a repeat exam for borderline students between 60-65.

Some of the testing members approached the challenges by emphasizing the assessment of papers outside the school (N=63). One of the testing members specified her concern on this issue as follows:

Geçen sene bir hocamız sınav kağıtlarını evine götürmüş, okumuş bize teslim etmiş. Aradan bir ay sonra bir öğrenci sınav sonucunun sistemde olmadığını ve bölüme geçmesi lazımkken, hazırlıkta görüldüğünü bildirdi. Hemen kırmızı alarm. Hocayı bulduk çünkü paketler kaybolursa geri dönüşümüz yok. Öğrenci mahkemeye verebilirdi, sorunu çözdük sonunda. Ama kağıt kaybetme lüksümüz yok. Bu yüzden kağıtların eve götürülmesi büyük bir sorun. Sonuçta kağıt kaybedilebilir, üzerine bir şeyler dökülebilir, yangında yanabilir ne bileyim her şey olabilir.

Last year, one of our instructors took the exam papers to his home, scored them, and handed them to us. A month later, a student reported that her exam result was not entered in the system and that she was displayed in the preparation system instead of the department. It was an immediate red alert. We found the instructor because if the packs are lost, we have no return. The student could take it to court, we finally solved the problem. But we can't afford to lose paper. Hence, taking the papers home is a big issue. After all, it is

possible they can lose papers, spill something on them, or they may get burnt in a fire, I don't know, anything can happen.

Finally, testing members rarely mentioned complicated score calculation (N=31). One of the testing members explained the issue in the following extract:

Bizim placementımız başta olmak üzere sınavlardaki hesaplamalar biraz karışık aklımızda biz de tutamıyoruz yani. Karıştırıyoruz, birbirimize soruyoruz. Sürekli çizelgeye bakmak zorunda kalıyoruz. Sürekli değişiyor. Her değiştiğinde tekrar rapor hazırlayıp üst yönetime bildiriyoruz.

The calculations in the exams, especially our placement, are a bit complicated, so even we can't keep them in our minds. We are confused and ask each other. We have to constantly look at the chart. It's constantly changing. Every time it changes, we prepare a report and submit it to the rectorate.

Out of nine challenges during evaluation, reliability and standardisation were top issues. Besides, most TUMs criticised the lack of well-developed rubric as they mentioned to have been employing randomly found ones even without adaptations or revisions. Another major issue was the instructors' failing to meet assessing deadlines, which was associated with their low work discipline. In the next section, challenges during post-evaluation will be explained.

4.1.3.4. Challenges during post-evaluation

This subcategory includes eight emerging codes which are shown in Table 4.15.

Table 4.15. Challenges during post-evaluation

Emerging Codes	Number of Codes	%
Over documentation	254	21,99
Unjustified objections to scores	189	16,36
Lack of archiving system	167	14,46
Lack of statistical analysis	152	13,16
Move up/down adaptations	134	11,60
Requirement for a fast announcement	104	9,01
Need for licenced programmes	87	7,53
Lack of prior documentation	68	5,89
Total	1.155	100

The most frequent anxiety in this subcategory was testing members' concerns about over-documentation (N=254). One testing member mentioned in the interview:

Sınavı hazırlamak bir yana, sınav öncesi ve sonrası kağıt işleri bizi çok yoruyor. Sınav öncesi tüm kağıtların fotokopi çekilmesi, zımbalanması, gruplanması, paketlenmesini yapıyoruz. Hocalara imza karşılığı teslim ediyoruz. Gözetmen listelerini hazırlıyoruz, hocaları ikna ediyoruz görev için. Sınav sonrası kağıtları alıyoruz, paketlemeleri yapıyoruz. İtirazları değerlendiriyoruz. Öğrencilerin kağıtlarını tek tek bulup, inceliyoruz ve raporunu yazıyoruz. En son arşivlemeyi yapıyoruz. Bunlar aklımıza ilk gelenler, anlatırken bile yoruldum. Ofiste çalışırken SAW filmi gibiyiz.

Aside from preparing the exam, paperwork before and after the exam makes us tired. We do photocopying, stapling, grouping, and packaging of all papers before the exam. We deliver them to the instructors with signatures. We prepare the supervisor lists and convince the instructors for the task. We take the papers after the exam and do the packaging. We deal with objections. We find the papers of the students one by one, examine them and write a report. We do the final archiving. These are the first things that come to our minds. I'm tired of even telling you about our workload. It is like the movie SAW when we work in the office.

The challenges related to students' unjustified objections to scores (N=189) comprise the second most frequently used type in this subcategory in the interviews. This is one of the main challenges of testing members as indicated in the following extracts:

Öğrenciler sınav sonuçlarına itiraz edebiliyorlar, tabii ki resmi itiraz yoluna gidiyorlar. Bu durumda, itirazda bulunan öğrencilerin kağıtları için tekrar bir değerlendirme yapılıyor, 3-4 gün içinde. İlk puanlama yapanlardan farklı, bir komite oluşturuluyor, 3-4 kişilik. Ancak sadece maddi hatalara bakılıyor. Diğer bir deyişle öğrenci dilekçe verdiğinde maddi hataya bakılıyor mesela içerik olarak writing yeniden okunmuyor, sınav tekrar değerlendirilmiyor. Onların onayı ve incelemesiyle sonuçlar açıklanıyor. Öğrencinin itiraz ettiği not güncelleniyor gerektiği takdirde. Genelde öğrencinin aleyhine gerçekleşiyor. Şüphesiz öğretmenlerimiz bol bol puan veriyorlar, mümkün olduğu kadar destek olacak şekilde vermeye çalışıyorlar. Üçüncü göz onları muhakkak kesiyor, biçiyor. Beklediklerinden düşük sonuçlarla karşılaşabiliyorlar açıkçası.

Students can object to exam results. Of course, they present their objections officially. In this case, a re-evaluation is made within 3-4 days for the papers of the students who objected. A committee of 3-4 people different from the instructors involved in the first scoring is formed. However, only technical errors or miscalculations are checked. In other words, when students submit their objection letter, technical errors or miscalculations are

checked. However, let us take writing, its content is not re-assessed, and the exam is not re-evaluated. With their approval and check, the results are announced. The grade objected to by the student is updated, if necessary. The results are usually to students' disadvantage. Undoubtedly, our instructors give more points while trying to give as much support as possible. The 3rd eye definitely cuts their scores. Obviously, they may get lower results than they expect.

Öğrencilerden gelen şikayetler çok ilginç, bazıları da klasik. Bir hoca yüksek vermiş, diğeri düşük vermiş. A sınıfındakilerin çoğu yüksek almış, B sınıfındakiler ise düşük almış. Kulaktan dolma, öğrenciler arasındaki dedikodular. Örneğin gramerde not kırıldığı için itiraz ediyor bir öğrenci. Yanlış olan cevabını bile doğru olduğunu düşünüyor ve yeniden okunmasını istiyor öğrenciler. Ya da writinglerinin tekrar değerlendirilmesini istiyorlar. Mesela geçen dönem, proficiency yaptığımızda 120 tane itiraz geldi. Üç günde 120 itirazın kağıtlarını arşivden bulduk, inceledik ve cevap hazırladık.

We receive very interesting complaints from students while some of them are classic. One instructor gave high while the other gave low. Most of the students in Class A got higher, while those in Class B got lower. Those are gossip among the students. For example, a student objected because she got a lower score in grammar. Students think that even their wrong answers are correct and want their papers to be scored again. Or they want their writing re-evaluated. For example, the last term, when we made proficiency, 120 objections were received. In three days, we found the papers of 120 objections from the archive, analysed them, and prepared an answer for each.

Another point was mentioned by a director:

Düşük alan öğrenci artık bizi şaşırtmıyor. Geçen sene geçen öğrenci ofise gelerek diyor ki B ile geçmişim ben niye A ile geçmemişim diye itiraz ediyor. Hani sırf kalan öğrenci değil. Kalan öğrenciler zaten hepsine, her şeye itiraz ediyor. Writingden mesela baya F, D veya NA almış yani çok düşük notlar almış öğrenciler bile yine de itiraz etmiş, yani şu kadar essay yazmış. Son örnekte, kağıdını bulduk, tekrar inceleyelim dedik ama gördük ki sadece instructionları yazarak paragraf yazmaya çalışmış.

The students getting low scores no longer surprise us. Last year, a student who passed the exam came to the office and said he got a B and asked why he didn't pass with an A. Not only do the students having failed make objections. The students who have failed already make objections to everything. Even the students who got F, D, or NA from writing, that is, very low scores, still objected, claiming they wrote good essays. In the last example, we found his paper, and examined it again, but it was interesting that he tried to write a paragraph by just using the words in the instructions.

A testing member drew the attention to another side of the issue:

İtirazların en sorunlu noktası belki de öğrenciler artık korkusuz. Hak aramak önemli, öğrenci itiraz etsin, hakkını arasın. Lakin bazı öğrencilerimiz var, hem haksız ama bir o kadar saldırgan. Öğrenci kalınca önce itiraz ediyor, notu değişmezse mahkemeye gideceğim diye önce şantaj ediyor, sonra şansımı bir deneyeyim diye mahkemeye başvuruyor. Mahkeme sürecinde öğrencilerin kağıdına ihtiyaç duyuluyor. Gidip tık şöyle elinizle koymuş gibi bulamıyoruz. Tek başına hocaların yapması zor olabiliyor, bir yandan ders anlatacak. Bu süreçteki işler bile ölçme ofisinin sorumluluğunda olunca, o kadar işin gücün arasında bir de onunla uğraşamıyorlar. Arşiv bile büyük bir şey aslında yani iş kalemi açısından. Hazırlamadan tutun dosyalamaya kadar bayağı bir yük. Üstüne bir de mahkemeye verilirse geri dönüp bakılıyor. Oldu daha geçenlerde bir mahkememiz vardı. Öğrenci hem notuna hem de geçme kalmaya itiraz ediyor. Final notuna galiba. Mahkeme bilirkişi atıyor, sanırım bir de bizden en son sınavda yüksek almış bir öğrencinin notunu ve kağıdını ve çok düşük almış bir öğrencinin notunu ve kağıdını istediler. Ayrıca mahkeme bizden bir de bu öğrencinin notunu ve kağıdını yani üç tane kağıt gitti mahkemeye.

Perhaps the most problematic point of the objections is that students are now fearless. It is important to seek justice, that is the students should object, and seek their rights. However, we have some students who are both unfair and aggressive. When some students fail, they object first, blackmail us that they will go to court if their scores do not change. Then they apply to the court in order to try their luck. Students' papers are requested during the court process. We can't find it immediately. It can be difficult for the instructors to do it alone while teaching and preparing tests. It is unfair that the work in this process is under the responsibility of testing offices as we cannot deal with it in the midst of all our work. Even the archive is a big thing in terms of workload. It's quite a burden from preparation to the filing. Besides, if it is taken to court, we need to make a search back. We had a court issue recently. A student objected to both his grade and passing score line. I guess it was about the final score. The court appoints an expert. I think they demanded us to submit the scores and papers of a student who got high and those of a student who got very low. In addition, the court also asked for the relevant student's score and paper, that is, three papers were sent to the court.

Lack of archiving system (N=167) has the third highest frequency among the challenges in this subcategory testing members face when codes are computed. Some testing members elaborated on the issue as follows:

Arşivleme konusunda, 1-yer sorunu, 2-ekstra iş yükü çıkmış oluyor. Arşiv ile ilgili çalışmalar ilk CEA akreditasyona gireceğimiz dönemde başladı. Öncesinde düzenli bir arşivimiz yoktu o güne kadar. Sonra yönetimi akreditasyon için ikna edemedik, sonrasında arşiv çalışmalarımız yarıda kaldı.

In terms of archiving, we can mention 1-space problem, 2-extra workload. Studies on archiving started when we began CEA accreditation process. Before that, we did not have an appropriate archive system. Later we could not convince the rectorate for accreditation, then our archive work was left aside.

Zaman zaman arşivleyen arkadaşlar, talep edildiğinde bulamayabiliyor sınavları, nasıl oluyor bilmiyorum. Hocamızın teslim ettiği ortada ama bir şekilde zarf bulunamayabiliyor, bir şekilde öğrencinin kağıdında bir eksik çıkabiliyor.

From time to time, our members who archive papers may not be able to find the relevant exam document when requested. I don't know how they can't find it. It is obvious that our instructor has submitted the envelope, but somehow the envelope cannot be found. Somehow the student's paper may be missing.

Aşağıdaki arşive kağıtlar gittikten sonra inşallah hiç gerekmez diye dua ediyoruz. Çünkü oradan bir şey bulmak herhalde azap olur diye düşünüyorum. Geriye dönmek oldukça zor. Sınavları paketlemek için bize çöp torbası veriyorlar. Siyah bir çöp torbasına koyuyoruz. Üstüne de hangi sınavın olduğunu yazıyoruz. Sonra görevli geliyor, beraber arşive gidiyoruz, görevli söylene söylene onları kurbanlık getirir gibi yerleştiriyor. “Siz zayıfsınız siz daha iyi girersiniz” diyor, bir de onla bir tartışıyoruz. Çöp torbası zaten ceset gibi. Ya karışsa o torbalar çöplerle. Türkçe yazıyor ki diğer çalışanlar anlasın diye. Çöp poşetine koymamızın sebebi onu daha korunaklı bir hale getirmekmiş. Ama iki gün sonra ben arşive insem, o poşeti nasıl açacağım, yırtmam lazım. Tabi bilgisayarda notlar var ama yazılı doküman gerekirse, yandık yani.

After we carry the papers to the archive down, we pray that we do not need them. I think it would be a torment to find something there. It is quite difficult to go back. They give us garbage bags to pack the exams. We put them in a black garbage bag. We also write the information about exams on it. When staff comes, we go to the archive together. The officer places them as if they are sheep to the slaughter. He says, “You are thin, you can fit better,” and then we argue with him. Garbage bag is like a corpse. What if those bags get confused with real garbage bags? We write pack info in Turkish so that other employees can understand it. The reason we put them in garbage bags is to ensure their protection. But if I go down to the archive after two days, how will I open that bag, I have to tear it. Of course, there are grades on the computer, but if a written document is needed, we are in a serious situation.

Another member explained further:

Mesela geçen sene ilginç bir olay yaşadık. Altı sene önce burada muafiyet sınavına girmiş kalmış ama geçtiğini zannediyor. Bölümüne devam etmiş, tüm derslerini bitirmiş, mezun

olacak. Bölümün ve fakültenin gözünden kaçmış. O zamanda hani %30-50 İngilizceler falan vardı. Bölümden diploma alacağını zannederken karşısına bir çıkıyor ki hazırlıktan kalmış. Öğrenci “ben geçtim” diye itirazda bulunuyor, “hayır kalmadım ben” diyor. Fakülte bize resmi yazıyla durumunu soruyor. Önce bilgisayardan notunu bulduk, gönderdik. Öğrenci kağıdının tekrar okunmasını talep ederek mahkemeye gideceğini söyledi. Kağıdını aradık ama uzun süre bulamadık. Rektörlükten yazı yazdılar, olay büyüdü. Sonunda yanlış packin içinde bulduk. Hepimiz istim üstündeydik.

For example, we had an interesting event last year. A student took the exemption exam here six years ago. Assuming that he passed, he continued his education at the faculty and completed all the courses waiting for graduation. It was overlooked by the department and the faculty. At that time, you know, there were 30-50% English. While he applied for a diploma from the department, it was realised that he had failed at preparatory school. The student claimed "I passed" and "no, I did not fail." The faculty inquired about his status in an official letter. First, we found the score on the computer and sent it. He said that he would apply to the court, demanding that his paper be scored again. We searched for his paper but couldn't find it for a long time. The rectorate wrote an official letter about the case. We finally found it in the wrong pack. We were at alert.

As a number of participants indicated in the interviews lack of statistical analysis of the exams (N=152). One of the testing members explained her concerns as follows:

İstatistiksel analiz için donanımlı teknik bilgiye sahip uzmanımız yok. Bunun için vakit problem de var. Hizmet içi veya hizmet öncesi eğitim eksikliğimiz de var. Program deseniz çok pahalı, öyle bir bütçemiz yok.

We do not have experts equipped with technical knowledge for statistical analysis. There is also the problem of time. We also lack in-service or pre-service training. The programmes are very expensive, but we do not have such a budget.

Another major concern of the directors and testing members related to national challenges is moving up/down adaptations (N=134). As stated by a testing member:

Seviye belirleme sınavlarına gelmeyen öğrenciler dersler başladıktan birkaç hafta sonra seviyelerinden memnun olmadıklarını ve bir üst seviyeye geçmek istediklerini belirtiyorlar. Okulun ilk iki ayı sürekli seviye değişikliği ile ilgili dilekçe ve talep geliyor. Seviyeler arasında değişiklik yaptığımız öğrencilerin bu defa gittikleri seviyede adaptasyon sorunu yaşadıkları gözlenmektedir.

Students who do not take placement exams state a few weeks after the classes start that they are not satisfied with their current level and want to move to a higher level. In the first two months, there were a number of demands and requests for level changes. It is observed that the students we changed between the levels had adaptation problems at the new level.

One of the sources of anxiety related to testing processes was testing members' stress about the requirement for a fast announcement (N=104). One testing member expressed her concerns:

Özellikle finalde ve proficiency sınavlarında çok kısa bir süremiz kalıyor, 2-3 gün içerisinde bütün değerlendirme işlemlerini tamamlamamız bekleniyor. Açık uçlu soruların standardizasyonu, toplantılar, cevap anahtarının değişmesi, değerlendiricilerin sorumluluklarına uymamaları gibi sorunlarla boğuşurken bir de zaman faktörü devreye girince işimiz çıkmaza giriyor. Bölümde hoca geç açıkladığında sorun olmuyor ama bizde hemen telefon gelir. Üst yönetimin buradaki süreçleri bilmemeleri, işlerin ne kadar zaman aldığını bilmemeleri yüzünden sınav bitince neredeyse açıklamamız isteniyor.

We have a very short time, especially in the final and proficiency exams. We are expected to complete all the evaluation procedures within 2-3 days. While we are struggling with problems such as the evaluation of open-ended questions, standardisation meetings, changes in answer keys, and the instructors not complying with their responsibilities, our work gets stuck when we are out of time. There is no problem in the department when a professor announces the results late, but we get a call right away. We are asked to announce whenever the exam is over. That is why the rectorate does not know the processes here and does not know how long the work takes.

Another challenge experienced by testing members was reported as the need for licenced programmes (N=87). The following extract exemplifies how a testing member was affected:

2017 yılında okulumuzda SPSS üyeliği vardı. Basit düzeyde sınavlar arasında istatistiki karşılaştırmalar yapıyorduk. İyi de oluyordu. Ancak SPSS'in süresi geçti, yeni üyelik 2500 dolar olduğundan yenilenmedi. SPSS in cracki çok fazla kullanılıyor bazı yerlerde, ama biz korsan program kullanmak istemedik. Bu yıl bir arkadaşımız ile yüksek lisans dersinden edindiği bilgileri kullanarak soru havuzu oluşturmak istedik. Tabi biraz istatistik çalışmak gerekiyor. Biz o çalışmaya başladığımızda bir software'e ihtiyaç duyduk. Okulumuz onu birazcık pahalı buldu almadı. Kullanılan sorunun nerede kullanıldığını görmemiz lazım, onun kaydını nasıl tutacağız gibi çalışmaları yoğunluğun arasında yaptık. E istediğimiz program alınmayınca biz de kendi çabalarımızla hani bir yere kadar uğraştık, ancak ikna edemedik. Aslında çok çalıştık iyi bir soru havuzumuz oluşsun diye hakikaten.

In 2017, our school had an SPSS membership. We used to conduct statistical comparisons simply between exams. It was helpful. However, when SPSS membership expired, the new membership was not renewed since it was \$2500. Spss in crack is used a lot in some places, but we did not want to use cracked software. This year, we wanted to form an item pool with a colleague by using the information he gained from his MA course. Of course, some

statistics are needed. When we started that work, we needed a software. Our school found it a bit expensive and did not buy it. We wanted to see where items were used, track them, etc. We continued the work a little more in the midst of the intensity. We tried our own efforts to a certain extent, but we could not convince our management. In fact, we worked very hard to create a good item pool.

A testing member shared similar views but also put emphasis on the lack of technical equipment and programmes:

Bir lisanslı programın yüklendiği bilgisayarımız vardı, ama on kişi idare ediyorduk o bilgisayarda. İnternet kafe kuyruğu gibiydi, birimiz kalkıyor, hemen diğeri oturuyordu. Güzel bir programdı. Audioları yaparken birleştirme yapıyorduk, araya instructionları koyuyoruz. Ama onun yapıldığı program tek bilgisayarda var on kişi onu kullanmak zorunda kalıyoruz. Birbirimize soruyoruz ne zaman yemeğe gidiyorsun, o arada ben yapayım falan diye. Aslında kurumsal üyelik olsa herkeste olması lazım.

We have a computer on which a licensed programme was installed, but we as ten instructors use it on one computer. It is like an internet cafe queue, one of us leaving and the other sitting down. It is a good programme. We use it to make audio, combine different audio, and add instructions inside. But the programme is installed on one computer, so ten people have to use it. We ask each other when the other is going to lunch; in the meantime, I can do something. In fact, we wish to have an institutional membership so that everyone could have it on their computers.

One final challenge in this subcategory was indicated by some testing members that anxiety appeared during the process when there was a lack of prior documentation (N=68). A testing member, for instance, reported:

Ofiste belge işi zaten çok fazla. Üstüne bizden önce bunca zaman burada dokümantasyon yapılmamış. Bir de öyle bir sıkıntı var. Bir yerden başlayalım diyoruz. Ama bir yandan akan bir süreç var sınav yağmuru adeta, diğer yandan da geçmişteki belgeleri oluşturmak, dosyaları yapmak. Önceki dönemlerden bize miras kalan öyle bir şey olmadığı için ne yaptığımızı biz de bilmiyoruz.

Document work in the office is already too much. On top of that, there hasn't been any documentation here all this time before us. There is also such a problem. We say we start somewhere. But on the one hand, there is a flowing process, almost like a rain of exams, and on the other hand, creating past documents and files. We do not know what we are doing, as there is no such thing inherited from previous periods.

According to the results of the study, the highest-ranking challenge in this category was over-documentation, which made TUMs feel like an officer rather than an academician. In addition, most TUMs complained about unjustified objections and lack of proper archiving system, which was thought to increase the workload of documentation. Another major challenge was moving up or down adaptations due to latecomers and false placements. The other most striking problems included requirement for fast announcements, need for licenced programmes, and lack of prior documentation. The next subsection deals with general/other challenges.

4.1.4. General/other challenges

The last stated type of challenges is other/general challenges (N=2.649; 15,35%) encountered in testing processes at EPPs in higher education. This category consists of the challenges related with examination, language education, and time.

Table 4.16. *General/other challenges*

Sub-categories	Number of Codes	%
Challenges related with examination	1.199	45,26
Challenges related with language education	362	13,67
Challenges related with time	1.088	41,07
Total	2.649	100

As Table 4.16. presents, the highest-ranking subcategory belongs to challenges related with examination (N=1.199; 45,26%). The second highest-ranking subcategory includes challenges related with language education (N=362; 13,67%). Challenges related with time (N=1.088; 41,07%) are given in the last subcategory. The Figure 4.5. shows the sub-categories and codes of the general/other challenges.

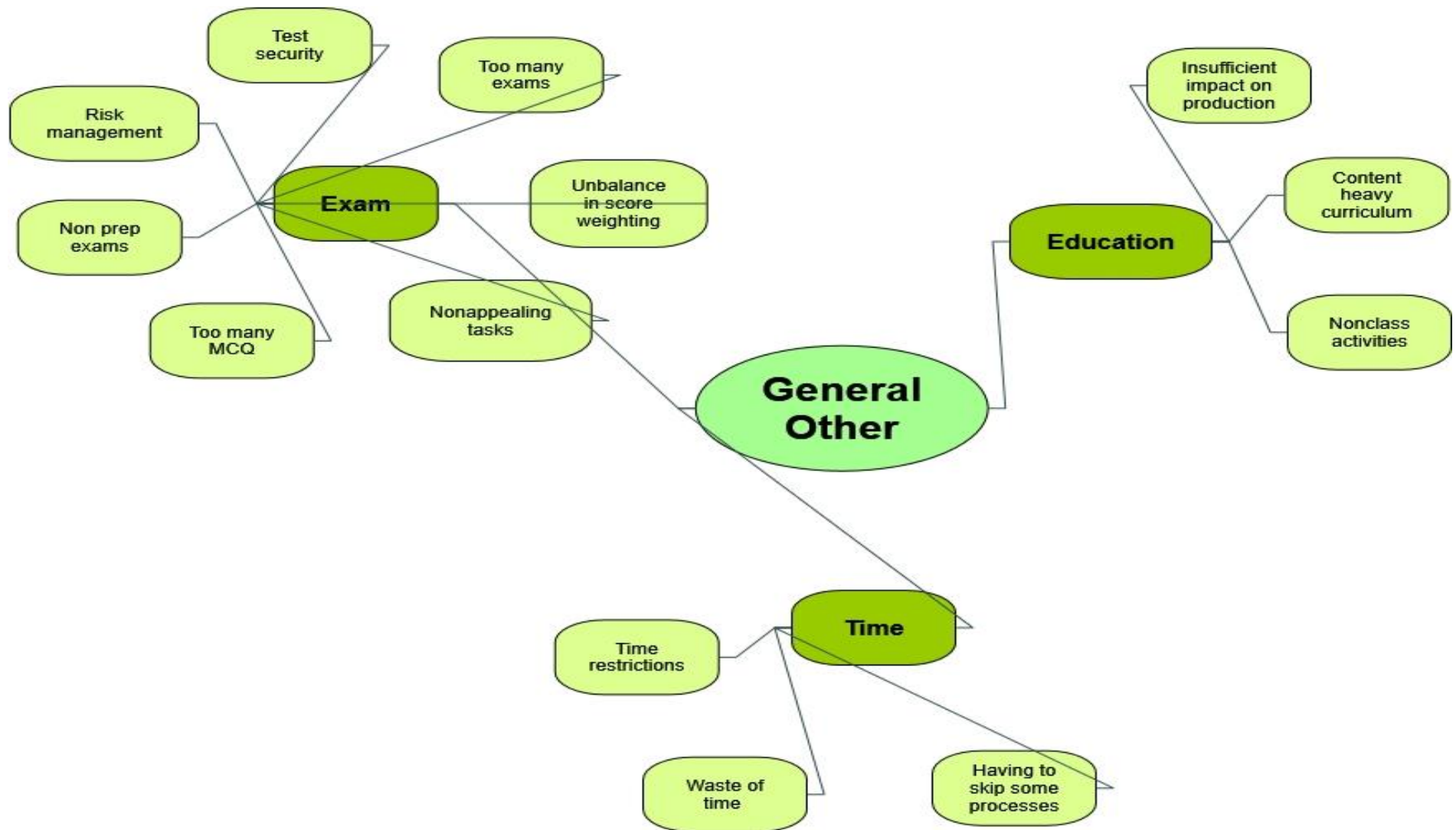


Figure 4.5. General/other challenges

The next subsections deal with the challenges related with examination, education, and time, explained respectively.

4.1.4.1. Challenges related with examination

This subcategory includes seven emerging codes as shown in Table 4.17.

Table 4.17. Challenges related with examination

Emerging Codes	Number of Codes	%
Too many exams	337	28,11
Test security	215	17,93
Risk management	185	15,43
Non- preparatory exams	180	15,01
Too many multiple-choice questions	131	10,93
Nonappealing tasks	89	7,42
Unbalance in score weighting	62	5,17
Total	1.199	100

Having too many exams (N=337) has the highest frequency among the challenges in this subcategory testing members face when codes are computed. A testing member explored as follows:

Yoğun bir sınav programımız var. Çok sık sınav yapıyoruz. En azından quiz filan yapıyoruz hem de her hafta. Vallahi yazık çocuklara her dersten her hafta quiz oluyorlar. Düşünebiliyor musunuz her dersten her hafta. Her levelin iki tane sınavı oluyor. Midterm olarak iki sınav, finalde iki sınav. Yedi görevli arkadaşımızdan birine her dönem iki ya da üç sınav görevi düşmüş oluyor. Yani kısa sürede iki ya da üç tane sınav hazırlamamız gerekiyor.

We have a busy exam schedule. We often carry out exams. At least we do quizzes every week. Poor children, they have quizzes from every lesson every week. Can you imagine? Every lesson, every week... Each level has two separate exams. Two exams as midterm, two exams as final. One of our seven colleagues is assigned to two or three exams each semester. In other words, we need to prepare two or three exams in a short time.

Another testing member added:

Çok fazla sınav hazırlıyoruz. Hazırlık programında proficiency, midtermler, finaller, quizler yapıyoruz. Bunların dışında UTE var mesela, mühendislik tamamlama var, yatay

geçişler var. Onlarda da yine proficiency gibi. Yabancı uyruklu öğrencilere bir şeyler oluyordu galiba, böyle ekstra sınavlarımız oluyor. Makeuplar var, sınava giremeyen öğrencileri bir daha makeup yapılıyor. Bir süreç devam ederken bir anda diğer sınav hazırlığı başlıyor. Boş bir takvimimiz yok. Sürekli bir hafta beş gün boyunca sürekli bir sınav sirkülasyonu var burada. Günde üç oturum bazen dört oturum olduğu oluyor. Bir sınav veriliyor, toplanıyor ondan sonra tekrardan başka bir gruba, başka bir fakülteye yapılıyor, öyle bir sirkülasyon var yani bir hafta gibi bir sürede. Zaman yetmiyor tabii ki evde, otobüste falan hazırlıyoruz.

We are preparing a lot of exams. We conduct proficiency, midterms, finals and quizzes in the preparatory programme. Apart from these, there is UTE, for example, one for engineering completion, there are transfer students. These are like proficiency. I guess we also prepare an exam for foreign students. We have such extra exams. There are make-up exams for the students who could not take the exam. While one process continues, preparations for the next exam begin. We don't have a blank calendar. There is a constant circulation of exams for five days a week. It happens that there are three sessions a day, sometimes four sessions. An exam is delivered, collected, and then another exam is given to another group, or to another faculty. There is such a circulation in one week. Of course, time is not enough, we prepare exams at home or on the bus.

The second most frequent anxiety in this subcategory was testing members' concerns about test security (N=215). One testing member stated in the interview:

Test security kısmı birazcık sıkıntılı. Hele proficiencylerde şöyle bir problem var her ne kadar technological tool kullanılmasını ve sınavda çıkanın fotoğraf çekmemesini de engellesek, şimdi soru bazında sınavlara girip böyle ezberleyen öğrenciler var. Birtakım kurslar bunu hedef edinmişler. İşte on tane öğrenci seçiliyor, her birine bir soruyu aklında tutsun diye görev veriliyor. Ya da eski sınavları toplu olarak da satın alıyorlar öğrencilerden. Bir şekilde sınavlarımız fotokopi merkezlerinde satılıyor, nerden geliyor oraya tam olarak bilemiyorum. Sınav bankamız büyük bir tehdit altında, ezberleyerek çıkıyorlar. Baya örgütlü çalışıyorlar yani. O kadar buna uğraşacaklarına ders çalışsalar ya. Geçen hafta alıcı gibi gittim, 2016-2017 yılı vardı ellerinde. Eski sınavlarda çıkmış soruları kitap şeklinde basmışlar, satıyorlar. Üniversite olarak önlem almadıkça, fotokopiciler, forum siteleri, kurslar para kazanıyor bu işten. Dışarıdaki kurslar sınav sorularını elde edebilmek için organize çalışmaya başlamış yani. Ki görüyoruz birtakım forumlarda, öğrenciler 5-10 tane soruyu paylaşmış. Adamlar tamamını elde edemese bile tehlikeli bir durum. Öğrencilerimizden haberdar olanlardan, sorumluluk sahibi olanlardan haber alabiliyoruz. Bazı forumlar için şikayette bulunduk ama sistemli, organize bir yapı var. Sonuç sadece sitenin kapanması, biri kapanıyor, b forumu açılıyor.

The test security part is a little troublesome. Especially in proficiency, there is a problem. Although we prevent the use of technological tools and students' taking photos in the exam,

now there are students who take exams and memorize the questions. Some private courses target this. They determine ten students, each given a task to keep a question in mind. They may also buy old exams from students. Somehow, our exams are sold in photocopy centres, I don't know exactly where they receive them. Our exam bank is under great threat, they leave after memorizing them. Therefore, they work very organized. I wish they'd study instead of trying these ways. Last week, I went like a customer, I saw the year 2016-2017. They have printed the old exams as a book and are selling them. Unless the university takes precautions, photocopy shop, forum sites and courses make money from this business. The courses outside started to work organized in order to get exam questions. We realised that in some forums as students shared 5-10 questions. It's a dangerous situation even if they can't get it all. We can hear from those who are aware of our students and those who are responsible. We have made a complaint on such forums, but there is a systematic and organized structure. The result is just the site closing, one is closed while forum B is established.

A testing member drew the attention to the separation of universities:

Öğrenci sayımız çok fazla. Üniversite ikiye bölünmeden önce kampüsler farklı yerlerde idi, hazırlık sınıflarımız da birbirinden uzak iki farklı kampüsteydi. Sınav dönemleri kağıtları taşıyorduk sürekli. Şimdi üniversitemiz bir kararla bölündü ama yeni üniversitede sistem kurulmadı. Bir hazırlık birimi yok Her sınavımızı böyle ilkel şartlarda şu gördüğünüz kolilerle xxx Üniversitesi'ne yollamaya çalışıyoruz. Ben mesela şahsen sınavın güvenliği açısından çok endişeleniyorum. Yani orada koli içinde giden sınava bir şey olsa, B planımız yok. Tabii ki orada yine güvenli yerlere bırakılıyor ama yani sonuçta sınavlarımız dışarıya çıkmış oluyor hocam yani sizin kontrolünüzden çıkmış oluyor. Öğrenci sayımızın bu kadar fazla ve okulumuzun tek kampüste olmayıp şehrin ayrı iki ucundaki kampüslerde olması bile sınav güvenliğini etkiliyor aslında, üniversitenin bölünmesi ek bir yük, ek bir zorluk getiriyor.

We have too many students. Before the university was split into two, the campuses were in different places, and our preparatory classes were on two different campuses far from each other. We used to carry exam papers. Now our university was divided by a decision, but the system was not established in the new university. There is no preparatory school. We are trying to send each of our exams to xxx University under such primitive conditions with the boxes. For example, I am very worried about the safety of the exam. If something happens to the exam carried in boxes, we don't have plan B. Of course, they are left in safe places, but our exams are out of our control. The fact that the number of our students is so high, that our school is not on one campus, and that our school is located on different campuses at two ends of the city affect exam security. In fact, the separation of the university has brought an additional burden and challenge.

The challenges related to risk management (N=185) comprise the third most frequently used type in this subcategory in the interviews. A testing member, for instance, considered:

Risk yönetimi olarak çok eksiğiz. Kendi çapımızda bazı önlemler alıyoruz. Son dakikaya genellikle bırakmıyoruz, sınavlarımızı basıp hazır görmek istiyoruz 1-2 gün öncesinden. Ama bilgisayar çökerse ne yaparız diye düşünmedik. Sistem çökmesi ile ilgili back up almıyoruz. Yedekleme yapmıyoruz. Bazı arkadaşlarımız mailine atıyor, ama müdürümüz gmailin tehlikeli olduğunu, bu yüzden gmail üzerinden sınav göndermiyoruz veya yedekleme yapmıyoruz. Her şey bu bilgisayara bağlı. Allah korusun diyoruz genelde. Bir kere oldu aslında 2016 yılında bilgisayar çökmüştü erişemedik sınava. Final sınavı mıydı neydi, ama sınavın yapılmasına az bir süre kalmıştı. Sistemde bir hata oldu. Neyse ki bilgi işlem sınavlarımızı kurtardı. Çok tırsmıştık.

We are insufficient in risk management. We are taking some precautions on our own. We usually do not try to complete our tasks at the last minute as we print out our exams in time and prepare them 1-2 days before. But we haven't considered what to do if the computer breaks down. Regarding system failure, we don't have a backup. We do not copy and store exams. Some of our friends send exams to their e-mails, but our manager warned us not to send exams or upload backups via drive since Gmail is dangerous. Everything is stored on this computer. We usually pray to God. It happened once, actually, in 2016, the computer system failed, so we couldn't access the exam. It was the final exam, and there was little time left before the exam. There was an error in the system. Fortunately, the computer department saved our exams. We were so terrified.

A testing member added some more details and elaborated as follows:

Geçen bir toplantıda öğrendim, akreditasyon sürecinden geçmekte olan üniversiteler bu konuyu gündemlerine taşıyorlarmış. Bizde maalesef yönetici düzeyinde veya testingde çalışan arkadaşlarımız düzeyinde farkındalık düzeyindeyiz. Konunun önemini anlatmaya çalıştım ama ne gerek var denildi. Aldığımız önlemler çok ilkel. Eve giderken elektrikten çekiyoruz her şeyi. Bu konuda çıkış kapısında yazılı ibare var. Şaka yapmıyorum, bu da seneler önce bir olay olmuş o yüzden. Prizde takılı bir alet mi varmış, telefon mu takılıymış, yıldırım düşmesine sebep olmuş. O olaydan sonra yangın çıkmasını diye elektrikte hiçbir şey bırakılmaması için yönetim bu yazıyı gönderdi.

I heard at a meeting last year that the universities going through the accreditation process are bringing this issue to their agenda. Unfortunately, we, both our managers and testing unit members, are at the level of awareness. I tried to explain the importance of the issue, but its necessity was questioned. The measures we take are very basic. We unplug everything when we go home. There is a sign about this on the exit door. I'm not kidding. This happened years ago. There was a device plugged into the socket or it was a phone

plugged in, which caused a lightning strike. After that incident, the administration sent a letter so that nothing should be left in the electricity so that a fire does not occur.

A testing member mocked at this issue:

Cumaya gittiğimde dua edeyim bununla ilgili. Bu da bir risk yönetimi sayılır.

I will pray about it when I go to the mosque on Friday. This can be considered risk management.

Furthermore, non-preparatory exams (N=180) are a noticeable concern of testing members. The following extract explains her feelings about a testing member:

Asıl sorumluluğumuz olan hazırlık okulumuzda dahi çok fazla sınav hazırlarken, bir de üstüne diğer sınavlar binince, neredeyse asıl sorumluluğumuza zaman kalmıyor. Biz, sadece hazırlık sınıfının sorularını yazmıyoruz. Erasmus sınavı oluyor, ne bileyim farklı okulların veya birimlerin öğrencilerine yönelik sınavlar oluyor, ekstra make up sınavları oluyor, bu durumda, gelmeyen öğrenciler için tekrar sınav yazılması gerekiyor. Yıl içinde hazırladığımız sınavların sayısı birazcık artıyor özellikle belli dönemlerde. Bu da bizi haliyle zorluyor. Saymaya çalışayım: Erasmus, Mevlana, yeterlilik, ek madde 1, ek madde 2, yatay geçiş, dikey geçiş, çift anadal... Haziranın sonuna kadar bir sınav süreci oluyor. Bir de zorunlu kredili sınavlarımız var.

Even in our preparatory school, which is our main responsibility, we prepare too many exams, but when the preparation of other exams is assigned to us, we may not have time for our main responsibility. That is, we don't just prepare the preparatory class questions. There are the Erasmus exams, exams for the students at different schools or units, extra make-up exams, In this case, a new exam has to be written for the students who could not take it. The number of exams we prepare during the year increases, especially in certain periods. This compels us as well. Let me count: exams for Erasmus, Mevlana, proficiency, additional item 1, additional item 2, horizontal transfer, vertical transfer, double/dual degree students... There is a busy exam period until the end of June. We also have exams for common mandatory courses.

A testing member shared the same perception and added some more details:

Erasmus ve Mevlana sınavlarına ek bir de ÇYDS Çukurova Üniversitesi'nin yüksek lisans programlarına başvuran öğrencilerin girdiği sınavı hazırlıyoruz. Bu sınavların takvimleri de değişiyor. Çünkü canları istediğinde Erasmus sınavı yapalım diyorlar. Bir ay sonra da Mevlana birimi arıyor sınav yapalım diyor. Zaman olarak uyuşmuyorlar, bir kere verilen sınavı tekrar da vermek istemiyoruz. Uygun olmuyor. Birinin başvurusu atıyorum martta diğerininki nisanda. Sınavı şubatta veya ekimde ortak yapalım dedik, öğrenci geçerli bir sınav puanı alsın ve istediği başvuruda kullansın diye önerdik. Ama Kabul etmediler.

Burada hiçbir şey yapmadığımızı düşündükleri için, 3-5 kişi bile olsa yeni sınav istiyorlar. Bu birimler Rektörlüğe bağlı ve birimlerin başındaki kişi de Rektöre yakın olduğundan emir gibi anında istiyorlar. Bu arada kendi sınavlarımızın telafileri de oluyor, onları da belirtelim. Geçen yıl böyle bir şey oldu sanıyorum raporlu öğrencilere yaptığımız makeup sınavı ile hazırlık dışına yaptığımız sınav çakıştı, üzücü tarafı dışarıya yaptığımız sınava öncelik vermek zorunda kaldık.

In addition to the exams for Erasmus and Mevlana, we also prepare an exam X-YDS for the students who apply to the graduate programmes of xxx University. The schedules of these exams also change. Whenever they want, they tell us to prepare an Erasmus exam. A month later, Mevlana calls the unit and requests a different exam. They do not set a mutual time. Besides, we do not want to give the same exam again. It is not suitable. One programme receives applications in March while the other in April. We decided to make a common exam in February or October so that a student getting a valid exam score may use it in any application. But they did not accept. Because they think we do nothing here, they want new exams, even for 3-5 students. Since these units are linked to the Rectorate and the head of the units is close to the Rector, they demand it immediately like an order. In the meantime, there are make-up exams for our own exams. I think something like this happened last year when our makeup exam was delivered to the students with a report that coincided with the exam outside of preparation. The sad thing was that we had to give priority to the exam for non-preparatory school.

Furthermore, the issue of too many MCQs (N=131) is a noticeable concern of testing members. As highlighted by a testing member:

Bundan iki sene önce yine daha genel, paraphrase these sentences tarzı etkinliklerimiz oluyordu. Skorlandırırken eşitlik sağlayamadığımız için farklı arkadaşlarımızın farklı yorumladıkları için ve sürekli şikayet ettiklerinden dolayı daha net cevaplar kullanmayı tercih ettik. Yani multiple-choice kullanmaya başladık, başta bazı kısımlarda bu tür soruları kullandık. Sonra herhalde rahatlık güzel geldi. Normalde hep eleştirdiğimiz klasik test usulü soruları çok sevdik, diğer sınavlarımıza da yaygınlaştırdık. Şimdi finalimiz dahi test formatında, çoktan seçmeli.

Two years ago, we had more general, questions like paraphrasing these sentences. We preferred to use clearer answers because we could not ensure equality when scoring. That is why different instructors made different comments and complained all the time. Thus, we started using multiple-choice. At first, we used this type of questions in some parts. Then I guess comfort caught us tight. We liked the classic test style. Although we normally criticize MCQ, we expanded them to our other exams. Now even our final is in test format, multiple-choice.

A different type of challenges in language testing that some of the participants indicated in the interview is nonappealing tasks to the students (N=89). A testing member, for instance, explained the issue as follows:

Sınavlarımızdaki en önemli sorunlardan biri soru yapılarımız ve konu başlıklarımız çok klasik ve eski. Dünyada, ülkemizde, bölgemizde veya şehrimizde her geçen gün küçük çaplı veya sansasyonel yeni olaylar olmakta, ancak biz hala üniversitede okul kurulduğunda oluşturulan konuları işliyoruz. O yıllarda iyi soru havuzu oluşturulmuş, ilginç. Zaman ilerliyor ancak biz eski hocalarımızın yaptıklarını bile güncelleyemiyoruz. Konular eski ve sıkıcı geliyor bu kuşaktaki çocuklara. Örneğin, Kadir İnanır ile bir paragraf var, çocuk kim olduğunu bilmiyor. Böyle olunca da çocuklar sınavlarda ilgisiz, üstünkörü cevaplar veriyorlar.

One of the most important problems in our exams is that our question structures and topics in the exams are very classic and old. In the world, in our country, in our region or our city, there are new small-scale or sensational events every day, but we are still dealing with the issues that were formed when the school was established at the university. A good pool of questions was created in those years, which is interesting. Time is ticking, but we can't even update what our previous instructors did. The topics seem old and boring to the students of this generation. For instance, there is a paragraph with Kadir İnanır, the student does not know who he is. Therefore, they give irrelevant, cursory answers in exams.

Finally, some testing members were anxious when there was unbalance in score weighting (N=62). A testing member elaborated:

Bizim okulda instinct'e bağlı bir puan skalası belirlenmiş. Production kısmı oldukça düşük. Geçme notunda quizlerin yeri azaltılmış. Hem sınavlar arasında hem de proficiency'deki beceriler arasında puanlama açısından denge yok. Hani bu puanlar neye göre belirlendi biri sorgulasa, bilmiyorum ben kendim justify edemem. Bu konuda bir rahatsızlık da yok. Tıpkı değişim isteyenler diye sorduğumuzda hep sistem eleştiriliyor, ancak bir alışkanlığı değiştirelim dediğimizde yeni adet çıkarmayın diyor eski hocalar.

In our school, a score scale was determined based on instinct. The production part is pretty low. The place of quizzes in the total grade has been reduced. There is no balance in terms of scoring between various exams and skills in proficiency. If someone questions me about how these scores were determined, I don't know and can't justify it myself. There is no discomfort about it. Just like when we ask those who want change, the system is always criticised, but when we suggest changing one habit, old instructors object and warn us not to introduce a new order.

In the category of the challenges related with examination, the most frequently mentioned challenge was the number of the exams at EPPs. It was claimed that they had difficulty taking measures on test security and risk management. Besides, non-preparatory exams such as Erasmus and Mevlana language exams, the use of multiple-choice questions, non-appealing tasks to the students, and unbalance in score weighting were other major challenges influencing the air in the TUs. The next subsection deals with challenges related with language education.

4.1.4.2. Challenges related with language education

This subcategory includes three emerging codes which are shown in Table 4.18.

Table 4.18. Challenges related with language education

Emerging Codes	Number of Codes	%
Insufficient impact on production	145	40,06
Content-heavy curriculum	122	33,70
No impact of non-class activities	95	26,24
Total	362	100

The challenges related to insufficient impact on production (N=145) comprise the most frequently used type in this subcategory in the interviews. A testing member explained as follows:

Sadece bizim üniversitemizde değil ülkemizdeki özellikle devlet üniversitelerinde ve gelişmesini tamamlayamamış vakıf üniversitelerinde production kısmı sorunludur. Tek bir sebebi yok belki, ancak bana göre başlıca sorun, YDS, YÖKDİL adı sürekli değişen merkezi sınavımızın ortaya çıkardığı washback effect. Üniversiteler için, dil eğitimi için bu sınavlar bir tehdittir, dinamittir. Öğrenciler bizim sınavlarımızda zorlandıklarında hemen bu sınavlara gidiyorlar, neden mi konuşma yok, yazma yok, üretime dönük ölçme yapılmıyor. Sonrasında üniversiteler olarak bizler de bir süre sonra cevaplama kısmı hassas bu production becerilerini sınavlarımızdan çıkarıyoruz. Production öğrencinin gözünde zaman kaybı görülüyor bu defa.

The production part is problematic in our country, not only in our university, but also in state universities and foundation universities that have not completed their development. There may not be a single reason, but for me, the main problem is the washback effect of our central exam, whose name is constantly changing, YDS, YÖKDİL. For universities and language education in general, these exams are a threat, dynamite. When students have difficulties in our exams, they immediately take such exams since there is no speaking, no

writing, no production-oriented measurement. Afterward, we, as universities, remove production skills, which are difficult to score, from our exams after a while. This time, production skills are considered a waste of time by the students.

A number of participants indicated in the interviews content-heavy curriculum (N=122). As explained by the participants:

Derslerimiz geleneksel tarzda gramer eğitime dayalı, böyle olunca tersten işliyoruz alında, yani otomatikman sınavlarımız da geleneksel. Dünyada yabancı dil edinimi ve sınavları 21. yüzyıl becerileri, iletişim becerileri, production becerilerine odaklanırken bizde hala neden sonra s takısı gelecek. Neden mi, çünkü müfredatta o konu var. Sınavlarımız müfredat odaklı, tıpkı sınıflardaki dil öğrenme ve öğretimi süreçleri gibi.

Our courses are based on traditional grammar education. We work backward, so automatically our exams are traditional, as well. While foreign language learning and exams in the world focus on 21st-century skills, communication skills, and production skills, we still try to ask if there is the suffix s after what? Why, because that subject is in the curriculum. Our exams are curriculum-oriented, just like language teaching and learning process in the classrooms.

Some of the testing members approached the challenges by emphasizing no impact of non-class activities (N=95). One of the testing members specified her concern on this issue as follows:

Listeningler zor gelebiliyor öğrencilere her ne kadar ilgili levele uygun olması için mümkün olduğunca uğraşılsa da. Normalde level problemi olmaması gerekiyor. Örneğin; elementary level da elementary level-1 listening dinlediklerinde dahi zorlanabiliyorlar. Bu konuyu araştırdığımızda, bunun genelde öğrencinin sadece ders içinde listening çalışması, ders dışında listening ile ilgili herhangi bir uygulama yapmamasıyla bağlantılı olduğunu görüyoruz. Benzer durum aslında writing için de geçerli. Öğrenciler paragraf oluşturamıyorlar. Hocalara bir anket uyguladık ve ödev yapma oranının çok düşük olduğunu fark ettik. Derste de bir sürü faktörden dolayı sınırlı geliyor veya hiç gelişmeyebiliyor becerileri.

Listening can be difficult for our students, even though we try as much as possible to fit the activities to the relevant level. Normally there shouldn't be a level problem. For example; even when they listen to the elementary level-1 listening at elementary level, they may have difficulty. When we examined this issue, we found out that it is generally related to the student's engagement in listening only in class and not doing any practice related to listening outside the class. The same is for writing, as well. Students cannot write paragraphs. We applied a survey to the instructors and noticed that the rate of doing

homework was very low. In the course, due to many factors, their skills develop limitedly or may not develop at all.

The results revealed that the highest-ranking challenge in this category was insufficient impact on productive skills due to washback effect and wrong EFL strategies in higher education. It was stated that TUMs were forced to remove speaking and writing parts due to several factors such as understaffing, difficulty in evaluation of productive skills, the number of the students who choose to receive YDS or YÖKDİL instead of learning English during one-year programme. Another major challenge was the content-heavy curriculum which is believed to exist even in K12. The last challenge was no impact of non-class activities. The next subsection deals with challenges related with time.

4.1.4.3. Challenges related with time

This subcategory includes four emerging codes which are shown in Table 4.19.

Table 4.19. Challenges related with time

Emerging Codes	Number of Codes	%
Time restrictions	425	39,06
Waste of time	378	34,74
Having to skip some processes	285	26,20
Total	1.088	100

The most frequent anxiety in this subcategory was testing members' concerns about time restrictions (N=425). One testing member mentioned in the interview:

Zaman sıkıntısı çok fazla oluyor. Sürekli bir sınav bitiyor, hemen ardından bir sonrakine başlıyoruz. O sınavı bitirme sevincini, başardık, rahatlayabiliriz deme fırsatımız olmuyor. O duyguyu yaşayamıyoruz. Son iki sene birazcık daha iyi, çünkü sınav sayısını dönem başında dörtten üçe indirdik. Aslında testing ile ilgili uygulama açısından yapmak istediğim çok şey var ancak zaman kısıtı nedeniyle risk alamıyorum. Sınavların güncellenmesi ile ilgili ihtiyacın farkındayız, fakat güncellemeye zaman olmaması bizi çaresiz bırakıyor.

We have time restrictions too much. Continuously, one exam ends, then another one starts afterward. We do not have the opportunity to feel that we have succeeded, we can relax, with the joy of completing an exam. We cannot experience that feeling. The last two years were a little bit better because we reduced the number of exams from four to three at the

beginning of the semester. Actually, there is a lot I want to do in terms of testing, but I can't take risks due to time constraints. We are aware of the need to update our exams, but the lack of time to update leaves us helpless.

As exemplified by a testing member:

Mesela XXX şu an zarf yazıyor, 100 tane zarf yazacak. Zamanı değerlendirmek açısından bir yandan zarf yazmaya devam etmesi lazım diğer yandan size cevap verecek. Bu işler yüzünden aslında bize ofiste yeterli zaman kalmıyor. Onu şu an yapmak zorunda, düşününce bile şu an komik geliyor. Diğer arkadaşımız YYY eski sınav kolisini arşive bırakıp gelecek, on dakika geç başlayacak görüşmeye. Şartlarımız böyle maalesef, bizler ne tez çalışabiliyoruz, ne de kendimizi bu konuda geliştirebiliyoruz. Zaman altındır derlerdi de burada çalışmaya başlayınca sözün önemini anladım.

For example, XXX is preparing envelopes, she will write 100 envelopes. In order to use time effectively, while she has to keep writing envelopes, she will answer you. Because of such work, we don't actually have enough time in the office. She has to do it now, it's funny when you think about it. Another colleague of ours, YYY will leave the box of previous exams in the archive and will start the meeting ten minutes late. Unfortunately, we have such conditions, under which we can neither write a thesis nor improve ourselves in this subject. They used to say that time is gold, but when I started working here, I realized the importance of the word.

Another challenge experienced by testing members was reported as a waste of time (N=378). The following extract exemplifies how testing members were affected:

Ofis zamanımız çok boş ve alakasız işlerle tükenmekte. Kağıt, fotokopi, zımba, zarf, kolileme, liste hazırlama, bilgileri tabloya girme, packleri hazırla, kağıtları dağıt, topla, say, eksikleri not al, ilgili hocayı bul, ne oldu diye sor, insanları gözetmenlik yapmaya ikna et, birbiriyle pair olmak istemeyenlere dikkat et, varsa dinleme için audio, writing için ek sayfa ayarla, sürekli hocaları uyar, öğrencileri uyar gibi görevleri saymaya çalıştım. Ancak henüz sınav hazırlama süreçlerine giremedim bile. Bu gibi sekretarya düzeyindeki işler bize zaman kaybettirmekte. Ofisteki enerjimizi sınav hazırlamaya veya sınav süreçlerini iyileştirme çalışmalarına haralayabilirsek ne güzel olurdu. Alakasız toplantılara çağırıyoruz zorla. Ama asıl görevimiz ne yazık ki eve kalıyor genelde, evdeki iş gücümüze dönüşüyor onlar bizim. Ofiste yaptığımız işler ise zaman kaybı.

We waste our office time on empty and irrelevant work. Let me try to list some irrelevant tasks such as paperwork, photocopy, stapler, envelope, packing, preparing a list, entering the information into tables, preparing the packs, distributing the papers, collecting them, counting, taking note of the missing ones, finding the relevant instructors, asking what happened, convincing people to accept proctoring, taking care of those who do not want to

be a pair with each other, prepare the audio for listening, additional pages for writing, constantly warning the instructors, the students. However, I have not even mentioned the exam preparation processes yet. Secretarial-level work like the above-listed wastes our time. It would be better if we could spend our energy in the office preparing exams or improving exam processes. We are forced to participate in irrelevant meetings. But our main duty, unfortunately, stays at home, they turn into our workforce at home. The work we do in the office is a waste of time.

The final challenge seems to be having to skip some processes (N=285). One testing member explained how she felt:

Zaman çok kısıtlı olduğu için her şeyi çok idealinde olması gerektiği gibi yapamıyoruz. Bazen bazı süreçleri atlamak zorunda kalıyoruz. Maalesef sadece olağan işleri yürütüyoruz. Hadi bir de ekstra şuna bakalım, diyemiyoruz. Sene içerisinde maalesef hiç öyle bir zamanımız olmuyor, daha da ideal yapsak diyemiyoruz, alanımız çok dar o anlamda. Vaktimiz yok. Hemen her şeyi ne kadar en kısa zamanda, nasıl yaparız diye uğraşıyoruz. Şunu değiştirelim, yeni bir şeyler deneyelim diyemiyoruz yani yoksa biliyoruz o da var, bu da var. Maalesef hepsi kalıyor, hatta önceden yaptığımız bazı çalışmalarını bile atlamak durumunda kalıyoruz. Sonuç öyle işte.

Since time is very limited, we cannot do everything as we should be ideally. Sometimes we have to skip some processes. Unfortunately, we are only doing normal work. We can't even consider looking at an extra thing. Unfortunately, we do not have such a time during the year. We know that we should make it even more ideal, our opportunities are limited in that sense. We don't have time. We are trying to find out how quickly we can do almost everything. We can't try to change one task and apply something new, otherwise, we know that there is much. Unfortunately, they all remain untouched. Even we have to skip some of the work we've done before. That's the situation.

Another point was mentioned by a director:

Örneğin, listening tracklerini hocalarımız sınıfta okuyor, aslında öncesinde kaydedebilir ve dinletebilir veya native speakera okutup kayıt alabilir, filtreleme yapıp kaliteli ürün oluşturabilir. Bazı geçerli metot ve teknikleri biliyoruz ancak bu tempoda, bu kadar az kişi ile, bu kadar fazla öğrenci ile, bu kadar kısa sürede, bu kadar fazla sınavı yetiştirdiklerine dua ediyoruz. Soru havuzu gibi böyle yeni çalışmalar yapılabilir, ama gerçekten eleman sayısını bayağı bir artırmak gerekir. Ekstra bir iş zaten onlar da, şu anlattığınız şeylerde bile bir kişinin yapacağı şeyler normal de değil açıkçası. Dediğim gibi insan sonuçta burada görev yapanlar da. Bir üye bir hafta hasta olduğunda, o bir hafta da tam sınav dönemine denk gelirse ne olur. Örneğin, XXX hocam, kadıncağız karnı burnunda geldi,

sınav haftasına denk geldiği için izin başlangıcı. Bir arkadaşımız hasta olduğunda serum falan aldı, izin almamak için. Hepsi olmasa da fedakarlar.

For example, our instructors read out the listening tracks in the classroom. In fact, they can record it beforehand and make the students listen, or have a native speaker read out, record and filter it to create a quality product. We are aware of some existing methods and techniques, but we pray that they complete so many exams in such a short time, with so few people, for so many students. Such new studies such as the question pool can be done, but it is really necessary to increase the number of members. They are extra jobs anyway. Even the tasks we describe are not normally logical for one person to do, frankly speaking. As I said, the members who work here are human, as well. What happens if a member is sick for one week and that one week coincides with the exam period? For instance, XXX, the poor lady, came to the office while she was pregnant. Normally she was normally on maternity leave, but since it coincided with exam week, she came. When a colleague of ours got sick, he took a serum, to not get permission. They are self-sacrificing, though not all.

A testing member drew the attention to another skill:

Sınavın içeriğini belirlerken, validity ve reliability sorunlarını en aza indirmek istiyoruz ama bir kapsam geçerliliği için örneğin kelime bilgisinin ölçülmesini alalım. Dört haftanın sonunda öğrencilere on tane soru soruyoruz ama öğrettiğimiz 100 tane kelime var. Normalde o kelimelerin minimum 30-40 gibi olması gerekiyor ki kapsam geçerliliği olsun ama zaman yetersizliğinden dolayı ideal şartları bilmemize rağmen es geçiyoruz. Tüm sınavı yaşadığımız sorunlardan dolayı iki saatte bitirmeye çalışıyoruz, sınırlı hareket alanımız var.

While determining the content of the exam, we want to regard the validity and reliability issues, but for scope validity, let's take, for example, testing of vocabulary knowledge. At the end of four weeks, we ask students ten words, but we teach 100 words. Normally, those words should be at least 30-40 so that there is validity in scope. Yet due to lack of time, we skip them even though we know the ideal conditions. We are trying to finish the whole exam in two hours due to the challenges we have experienced. We have a limited range of movement.

According to the results of the study, the highest-ranking challenge in this category was about time restrictions, which pushes TUMs to compete with time in every step with a dead-end cycle as one exam ends, the other begins. In addition, most TUMs complained about waste of time, which was thought to result from secretarial work and redundant duties. The last problem was having to skip some processes due to understaffing and heavy workload.

To sum up, in the first phase, it was aimed to identify the challenges in assessment practices encountered in TUs at EPPs. The interviews with both directors and TUMs helped the researcher have an in-depth analysis of the challenges. The interviews with TUMs were conducted separately from the ones with directors in order not to make them affected during the interviews. The findings of the study showed that the challenges varied from one university to another depending on the characteristics of the university, region, and other criteria. However, the aim of the study was not to identify how they varied but to include as many challenges as possible to draw a map of problems encountered in assessment practices at EPPs to have a look at the general picture.

The analysis of the study reveals that there are four main categories of challenges, namely systemic challenges, challenges with stakeholders, testing processes-related challenges, and other/general challenges. It was recognised that one category of the challenges seemed to affect the other category or categories. What is more, it was found that some codes in a category appeared to impact other codes as challenges. This shows they are connected to each other like the count-wheel of a clock, flowing as chain.

The first category was on systemic challenges, including three sub-categories, national, institutional and unit-level challenges. It was clear that national challenges with 13 codes were in a higher order in terms of their impact. Major three challenges in this subcategory included legal ground necessity, uncompensated extra work, and lack of job description. The second subcategory was about institutional challenges which consisted of 12 codes. This included lack of pre-service or in-service training, constant system change, and poor physical conditions as major three codes. The last subcategory was about unit-level challenges with nine codes. Major three challenges in this subcategory included excessive workload, being understaffed, and excessive TUM turnover.

The second category was on challenges with stakeholders, including four sub-categories, challenges with administrators, TUMs, instructors, and students. Major three challenges in challenges with administrators included *fait accompli*, administrator change, and administrators from outside the ELT field in addition to other three codes. The second subcategory was about challenges with TUMs which consisted of eight codes. This included emotional factors such as demotivation, burnout, and anxiety as

major three codes. Major three challenges in third subcategory, namely challenges with instructors included tense reactions, distractions, and irresponsibility in addition to other five codes. The last subcategory was about challenges with students with ten codes. Major three challenges in this subcategory included the number of students, low English level, and gap between students' levels.

The third category was on testing processes-related challenges, including four sub-categories, namely challenges during preparation, implementation, evaluation, and post-evaluation. Major three challenges in challenges during preparation included approximation to students' levels, difficulty in item preparation, and scarcity of resources in addition to other seven codes. The second subcategory was about challenges during implementation which consisted of seven codes. This included large-scale exams, technical issues, and inadequate proctoring as major three codes. Major three challenges in third subcategory, namely challenges during evaluation included reliability issues, difficulty in standardisation, and lack of well-developed rubric in addition to other six codes. The last subcategory was about challenges during post-evaluation with eight codes. Major three challenges in this subcategory included the over documentation, students' unjustified objections to exam scores, and lack of archiving system.

The last category was on other/general challenges, including three sub-categories, namely challenges with examination, language education, and time. It was clear that challenges with examination with seven codes were in a higher order in terms of their impact. Major three challenges in this subcategory included the number of exams, test security, and risk management. The second subcategory was about challenges with language education which consisted of three codes. This included insufficient impact on production, content-heavy curriculum, and no impact on non-class activities. The last subcategory was about challenges with time with three codes. The challenges in this subcategory included time restrictions, waste of time, and having to skip some processes.

These challenges identified in the first phase were employed as the basis of the following phase in the focus group interview with the experts to find feasible solutions. The next subsection explains the findings of the second phase.

4.2. Findings of the Second Phase

This study aims to seek the challenges experienced in language testing at EPPs and possible solutions accordingly. In this phase, it was attempted to find solution suggestions by field experts employing the Delphi method. It consisted of three rounds the first of which served for getting the experts to focus on drafting their solutions considering the challenges discovered in the first phase, but the last round was to ensure a common view and consensus on the solutions anonymously.

Analysis of the qualitative data obtained from the focus group meeting with seven field experts indicated three main categories as the solutions to be a remedy against the challenges in testing processes. The main categories are given in Figure 4.6.

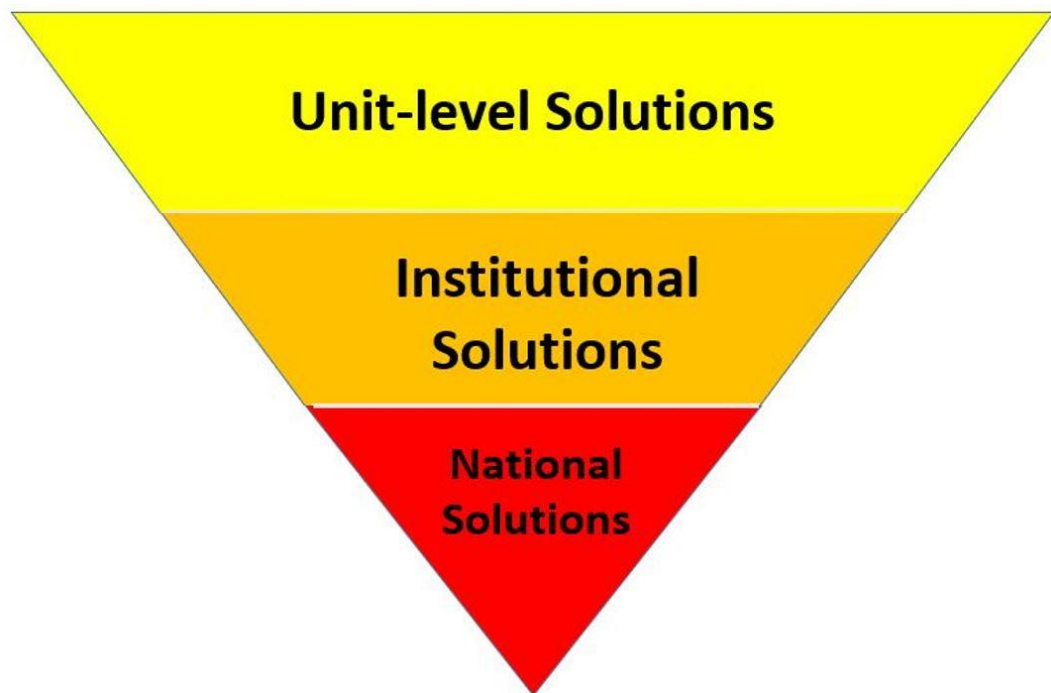


Figure 4.6. *Main categories of solutions*

As shown in Figure 4.6, the colors of the solutions categories go from red at the bottom to yellow at the top. This color differentiation was used to explain that national solutions were found to be most critical requirements to reach sustainable and effective results in the system. Following the national solutions, the institutional solution suggestions in orange were considered more important than unit-level ones. Hence, it could be inferred that without any improvement in national-level suggestions, whatever

measures TUMs attempt to take in TUs would be the tiny ship on the ocean against the huge waves. That is why they end up with the feeling burnout after a short time but have to continue to work unwillingly. Besides, the solutions were given similar to Maslow's Hierarchy, but in reversed position, which shows that although national-level solutions are fewer than unit-level solutions, they appear to impact the others as the main factor.

The remaining subsections explain each category and subcategory related to the solutions to the challenges in testing processes in detail based on suggestions that fall into each category. None of the categories existing in the literature were set as the basis for categorisation. Instead, upon the consensus among the experts, emerging codes that show similar characteristics were placed under certain categories in line with content analysis. The categories and solution suggestions are presented with extracts from the Focus Group interview with the field experts (See Appendix 6 for the list of suggestions in each category).

Analysis of the qualitative data obtained through a focus group interview with the experts indicated three main categories. The main categories are given in Table 4.20.

Table 4.20. *Main categories of solutions*

Categories	Number of Codes	%
National-level solutions	263	37,09
Institutional-level solutions	235	33,15
Unit-level solutions	211	29,76
Total	709	100

As Table 4.20. presents, the highest number of codes belongs to national-level solutions (N=263; 37,09%) while the second highest ranking category includes institutional-level solutions (N=235; 33,15%). Unit-level solutions (N=211; 29,76%) are given in the last category. The next subsections deal with solutions at the national, institutional, and unit levels, respectively.

4.2.1. National-level solutions

National-level solutions comprise the most frequently stated type of solutions to the challenges encountered in testing processes at EPPs in higher education. This category includes fifteen emerging codes which are shown in Table 4.21.

Table 4.21. *National-level solutions*

Emerging Codes	Number of Codes	%
Establishing the legal ground	29	11,03
Defining TUM title and job description	25	9,51
Delivering a national 4-skills-based exam	23	8,75
Leaving decision on exam equivalence to universities	21	7,98
Establishing HE Language Qualifications Framework	20	7,60
Ensuring mutual recognition	20	7,60
Providing financial incentives	19	7,22
Adding some criteria to Associate Professorship applications	18	6,84
Integrating exam preparation into YÖKSİS account	17	6,46
Including TUM into item writer pool	17	6,46
Including EPPs in university visit programmes	16	6,08
Table 4.21. (Continued) National-level solutions		
Conducting a national-level needs analysis	13	4,94
Determining bottom-up quality parameters	10	3,80
Regarding the capacity of EPPs for quotas	9	3,42
Establishing consortiums or body institutions	6	2,28
Total	263	100

The first issue accounts for the legislation necessities of TUs and the TUMs within EPPs. The most frequent solution in this category was the establishment of the legal ground of testing units. All of the experts firmly asserted that the authority council took this issue into its agenda. One expert mentioned in the focus group meeting:

“There is a serious human resource in our country who can improve themselves in this field. If the authority council realizes this issue and decides to take steps on necessary legal adaptations, we can prepare even international exams.”

Another expert explained in a more detailed way:

“YÖK held a workshop on this subject. I wish legal studies were carried out in line with that report. This can also be applied to other units at these schools.”

He continued to provide more information on the workshop that was held in 2017 with the participation of EPP directors at the council. It appears that the council is aware of the challenge, yet some other issues must have been prioritised. Within the

legal groundwork, it was also suggested that the title of testing unit member and job description be defined in the legal document. This is one of the main solutions as indicated in the following extract:

“If a legal study is to be carried out, a key item of the legal document is to determine the title of the people who will work in testing offices and define their job descriptions clearly. In this way, testing members may get more motivated and school administrators can take official actions more confidently both in terms of assignment and distribution of tasks.”

It seems that having to work in the testing office without a legal title leads to demotivation among testing members as this situation is seen to have made administrators take unofficial actions to maintain these offices. In addition, lack of job description is believed to result in ambiguous structure in these offices as there is no or limited definite task sharing and workflow in a number of EPPs.

The second issue is related with development of a comparable skill-based exam. The importance of a four-skills-based exam has the third highest frequency among the solutions in this category when codes are computed. It was proposed that ÖSYM should conduct a new format in language exams and prepare a four-skills-based exam at a national level instead of test-based ones. An expert elaborated on the issue as follows:

“I heard that some serious work has been done on this issue at ÖSYM, but no one among us knows at what stage it is. It can be a milestone study that will pave the way for language learning.”

She highlighted that it is obvious neither students nor academic staff is glad about the existing exams. According to the experts, until such a u-turn is made in language examination, the universities should be given freedom to decide on equivalence and recognition of YDS/YÖKDİL. As stated by an expert:

“We make a plan at the beginning of each semester and year. We determine the number of classes, student groups, and arrange instructor assignments accordingly. But suddenly a great number of students get a passing score from YÖKDİL because it is simple. Therefore, they do not attend the class anymore. Imagine there are only two students left in one class. How can the instructor lecture with two students? Thus, the equivalence of these exams should be left to universities.”

Agreeing with this idea, an expert who complained about the exit levels suggested that initially this freedom had better be strengthened with a national-level framework and the authority council should establish Higher Education Language Qualifications Framework in which exit levels need to be revised. One expert expressed her ideas about the issue:

“One of the reasons for the difference among universities is the lack of a common umbrella framework in our country; therefore, a framework can be established under the coordination of YÖK, as an authority institution, through workshops and meetings attended by relevant administrators and expert groups from universities.”

The experts took this issue one step further as such a framework at a national level could ensure mutual recognition among Turkish universities. One expert disagreed on the mutual recognition stating her anxieties:

“There are various universities from east to west, from developed to new, with various qualities. How can we accept a score from another university having different student types, different institutional cultures, different instructors, different curricula, especially different exam formats, etc.?”

However, at the end of the discussion, they all agreed that the national language framework has the potential to bring the universities setting the minimum standards and quality.

The third issue is about national incentives or initiatives for TUMs working in TUs. As most experts proposed, national-level incentives have a great place in motivating academicians in Türkiye. The experts who are currently working as school director at EPPs firmly asserted that financial incentives to be delivered by government authorities such as YÖK or TÜBİTAK would foster language instructors to volunteer to be working in these offices. One of the experts explained as follows:

“Financial incentives can be provided by YÖK by giving some credits to the testing members for the preparation of high-stake tests in academic incentive applications.”

Another expert shared similar views but also puts emphasis on Prof. Hasan Mandal's participation in the language school director's meeting held at Ankara Social Sciences University in 2017 when he was the Deputy President of YÖK. He explained further:

"Once, the current President of TÜBİTAK attended our directors' meeting when he was the Deputy President of YÖK in 2017. Before the meeting, a new official letter was sent from YÖK about academic positions within EPPs. In the first session, this article was discussed to lead to very different views and applications. However, in the second session, Prof. Mandal explained the real purpose of the letter. Finally, he stated that these schools have an important mission in THE system. Considering his approach to our problems in this example, I am sure he will be willing to offer help if our incentive proposal is sent to him via a suitable channel. In this way, TÜBİTAK can establish a new call towards language testing."

In addition, some criteria can be added to ÜAK Associate Professorship applications as a testing member may receive some extra points for working in these units. An expert, for instance, reported:

"In the past, those working in preparatory schools were not allowed to conduct graduate studies. Therefore, the number of people who had a Ph.D. degree was very few. However, as the restrictions were abolished, the number of graduate students increased gradually. One of the goals of our friends with doctorate degree is to pass associate professorship exam. In this case, in the application for associate professorship, giving additional points to those who serve in such offices will increase voluntariness."

Besides, exam preparation could be displayed at YÖKSİS accounts, which is also considered in reappointment process of academicians. The following extract exemplifies the suggestion of an expert:

"Generally speaking, the YÖK system offers two types of information about an academician. One consists of the training and certificates received while the other consists of academic products such as proceedings, articles and projects. Considering even certificates are added, large-scale exams can also be added to this system, which will be a tangible proof of their products for those who work here."

In this way, TUMs can be included in item writer pool by ÖSYM based on their success in exam preparation in their academic accounts. As highlighted by an expert:

"The ones writing exam items for ÖSYM earn extra income. However, it is not clear how and why some certain people are selected to write items. To make the system transparent

and inclusive, the selection can be done through an application call. Out of the applicants, success points shown in YÖK accounts might be used to list the applicants.”

Last but not least, it is suggested that YÖKAK may include EPPs in university visit programmes. It is believed that quality studies within the university seem to remain at surface or documentation level in terms of EPPs if they are not involved in the process and visits. One expert explained its importance by giving a warning:

“That the director of a preparatory school or a representative selected from the school takes an active role in quality studies can enable the school staff to get used to the concept of quality and adopt this notion, thereby increasing the internalisation of the quality culture among the staff. Otherwise, when the instructors and TUMs, having a lower level in university management are forced to prepare documents without establishing a quality mission, these vital studies turn into a formality while documentation can be seen as an extra burden and secretarial work for TUMs, who have extremely busy schedules. This situation may disrupt not only the school but also the university on the way to internationalisation.”

The last issue focuses on a bottom-up needs analysis of TUs. In the workshop, experts were in full agreement that needs analysis should be conducted at a national level. One expert expressed her suggestion:

“Academic studies are constantly being conducted about language testing on different subjects such as quality, type, and effects of exams because the subject is actually interesting in many ways. However, these studies are quite insufficient to address our problems completely. A national needs analysis similar to yours should be prepared so that the issue can be addressed with a holistic approach.”

Established in 2015, THEQC, in its basic documents, aims to ensure quality in the studies carried out in Turkish higher education institutions. In this way, it attempts to bring THE system to an advanced level in terms of international competitiveness. Therefore, quality parameters about these schools should be determined bottom-up, not top-down. One of the experts exemplified as follows:

“The presence of representatives from EPPs in the university visits of the committees selected by the Quality Board can create a positive basis for the decisions to be taken.”

The report published by the THEQC can also be sent to YÖK. In this way, these reports can be taken into account when student quotas and academic posts are allocated to universities by YÖK, which ensures that the views of EPPs on their capacity could be regarded by YÖK while delivering quotas and academic posts. An expert elaborated on this issue and stated that:

“YÖK sends more student quotas than demands of university. However, even the preparatory school, let alone the department, does not have enough academicians to meet this submission. Therefore, if YÖKAK reports are taken into account in both student quota and the appointment of academicians, more sustainable conditions will be provided.”

To ensure a more sustainable environment in EPPs, consortiums or body institutions, one state and one foundation, can be established considering geopolitical and institutional relevance. It was mentioned in the first phase that due to financial budget limitations of EPPs, TUMs were not able to take training on key issues; however, this suggestion might be a remedy until a regulation on the legal status of EPPs is maintained. The following extract would be enough to clarify the issue:

“Universities can also benefit from each other's experiences and methods. In fact, there may be such an implementation under the Anadolu project, because the nature of this project is to provide know-how from developed universities to newly established ones. In this way, universities can be grouped in terms of their practices in testing and plan joint activities to share good applications.”

According to the findings of the study, there are 15 codes as solutions in the first category, namely national solutions. The experts stressed the importance of legal studies, comparable skill-based exam, incentives and bottom-up needs analysis at national level. The first three codes included establishing the legal ground of the units, defining TUM title and job description, and delivering a national four-skill-based exam by ÖSYM. Besides, it was suggested that universities should be given freedom to accept which exams for exemption. However, the authority councils' supports were expected in the design of a framework, mutual recognition, and incentives. Furthermore, a national-level needs analysis was suggested, which could be used in quality studies, as well.

4.2.2. Institutional-level solutions

Institutional-level solutions comprise the second most frequently stated type of solutions to the challenges encountered in testing processes at EPPs in higher education. This category includes twenty emerging codes as shown in Table 4.22.

Table 4.22. *Institutional-level solutions*

Emerging Codes	Number of Codes	%
Working on digitalisation of exams	19	8,09
Giving priority to training	18	7,66
Prioritising TUM in assignments	17	7,23
Collaborating with international organisations	17	7,23
Improving physical conditions	16	6,81
Giving technical and sanitation staff	16	6,81
Providing licenced programmes needed	15	6,38
Integrating EPPs into automation system	14	5,96
Dividing preparatory education into three terms	13	5,53
Adopting level-based system	13	5,53
Appointing the EPP director as vice rector or advisor	12	5,11
Requiring service in one of the units at EPPs	12	5,11
Taking comments from TUs on key issues	11	4,68
Ensuring coordination among all units	10	4,26
Providing orientation to students	8	3,40
Fostering the continuity of professional development	7	2,98
Preparing a common exam calendar	6	2,55
Table 4.22. (Continued) <i>Institutional-level solutions</i>		
Maintaining relations of TUM-Instructors-Students at management level	5	2,13
Authorising TU for complaints/objections	4	1,70
Taking sanction decisions	2	0,85
Total	235	100

The first issue accounts for training on digitalisation. Digitalisation of exams has the highest frequency among the solutions in this category when codes are computed. All the experts were in full agreement with the necessity of digital exams due to the COVID-19 pandemic. An expert, for instance, explained the issue as follows:

“Pandemic has now shown the world the power of the virtual environment. As our country, we must be prepared for this. Our schools should make an investment in the digitalisation of language exams.”

However, they also confirmed that there were some drawbacks of online testing as one of the experts touched on the difficulty of online testing:

“It is very difficult to control, we make them open the camera, but we do not know if there is a mechanism behind the camera. It is not clear whether we can impose sanctions on anyone who does not want to open the camera. It is very open to abuse.”

Nevertheless, she asserted that digitalisation is at the door, and suggested that we need to leave the door open. It was admitted in the interview that testing members seem to lack knowledge of language testing since they may not receive pre-service/in-service training. They recommended that testing members should be provided pre-service/in-service training on both physical and online test preparation. The following extract exemplifies the issue:

“If testing members in these units receive pre-service and in-service training, they may be open to innovations. Only then can we get digitalised and produce more reliable digital exams.”

It appears that the delivery of digital exams can be worse for testing members since they may not be familiar with online platforms. In this regard, the experts proposed that testing members could be given priority in national and international assignments so they may attend seminars, conferences, or workshops. One of the experts touched on the potential opportunities:

“Since our resources are very limited, we ask publishing houses for training on certain subjects. They agree to organize workshops and bring experts if we buy books, that is, we are obliged to work with them. In fact, it would be beneficial if the members working in these units were given priority in assignments within national and international mobility programmes or conferences.”

Further, he gave staff exchange under Erasmus as an example. In addition, collaboration can be made with international organisations such as Fulbright or the British Council for developing EPPs, including Native speaker appointments. Experts stressed that since these organisations aimed to transfer the UK's expertise in education and especially English language to the countries they were registered, they might work directly with EPPs and TUMs to help them gain new skills, confidence, and connections to transform their work and shape a better testing system in partnership with universities in the UK. One of the experts explained how such organisations might contribute to the testing system at EPPs in the following extract:

“Before the pandemic, several international Fulbright instructors were working at EPPs, which provided motivation for both students and Turkish instructors. Another benefit of them is their contribution to the feedback cycle as a native speaker in the exams. In addition to COVID-19, restrictions by YÖK on foreign instructors influenced such appointments. In order to ease the shortage of native speakers, Fulbright instructors, in consultation with YÖK, can be counted out of the foreign instructor quota. In addition, the British Council can organize online seminars in England with the participation of qualified experts in testing at the University of Sheffield.”

The second issue is related with the physical and technical needs of EPPs. The fifth most frequent solution in this category was about physical conditions of testing units. It was noted that testing members should have separate rooms for working, archiving, and other needs. Having to work with more than four people in a small room or the basement with insects or without a window and air conditioner might be challenging. These offices should be provided necessary materials such as shelves, desks, computers, copy machines, etc. An expert explained as follows:

“Clean, spacious, and adequately sized room(s) should be given to these units. The cleaning of the rooms should be done as in other faculties or units. In addition to other technical equipment such as computers, hard disks, flashes, printers, scanners, stationery equipment, furniture, and inventory stock should be supplied as in the room of the school secretary.”

In addition to physical equipment support, one of the most commonly stated solutions was the need for technical staff and sanitation staff. Appointment of especially a technical staff seems crucial in order to avoid problems in the processes such as employing statistical programmes and delivering, scoring exams on digital platforms. One expert expressed her suggestion in these words:

“Unless our colleagues we assign in the testing unit have special training or self-study, their knowledge of statistics and software programmes is limited. Therefore, in these units, there should be a technical staff who searches, examines, employs technological devices, different software programmes or applications, and guides the testing members when necessary.”

Besides, considering the unity within EPPs, it is advised that these units be cleaned regularly by the sanitation staff of the school. As clarified by an expert:

“Sanitation staff should not only clean the classrooms, but they can also help these units in cleaning, carrying boxes, arranging bookshelves or other shelves when necessary.”

Assignment of assisting staff seems to increase the quality of exams and performance of TUMs. Besides, it was highly suggested by the experts that licenced programmes needed be provided. In the focus group interview, the need for software applications in TUs was emphasized by the experts who mentioned that various applications could be used while preparing questions, conducting the scoring, or analysing the results in TUs. One expert clarified their importance:

“We can use Maxadudit to combine audio; likewise, we can use SPSS for the analysis of the results. Although these programmes have free either temporary or full versions, you cannot license them without a premium membership. In this case, they work like a demo version. On the other hand, they can be used illegally through crack save, but it is unethical to install a programme on a computer in a government institution. In fact, these programmes can be added to the list of requirements taken from the units at the beginning of each year.”

It is also important that EPPs must be integrated into the automation information systems. The experts commented on this challenge as a common issue explaining that each university used an automation system, either creating its own system or through the purchase of services from firms. However, it was criticised that these systems would

not take EPPs into consideration in terms of announcements, exams, and calendar. An expert elaborated on the issue:

“University rectorate should inform Proliz system representatives of the different practices and problems of EPPs. For example, the system needs to adapt to different sub-scores of quizzes administered by EPPs. Furthermore, a common exam calendar should be prepared for all student groups in different units.”

The third issue is about structural priorities at EPPs. The experts indicated that preparatory education should be divided into three terms claiming that it appears difficult to achieve B1 or B2 targets in two semesters considering the starting levels of incoming students. An expert, for instance, put an emphasis on the inefficiency of EPPs in two terms and stated:

“Whether level-based, semester-based or even year-based system is applied, duration of two terms, that is eight months, is not enough for achieving learning outcomes and goals we want. Extending preparatory education into three terms, including summer, will increase the efficiency of the system and enhance better language acquisition, desire, and awareness of the students.”

In addition to extending the education period, according to the comments of the experts, a level-based system should be adopted. Acknowledging the difficulties of the system, an expert explained in the focus group meeting:

“The level-based system may have difficulties, but it helps students focus more on language skills. One of the most important features of this system is that the exams can be more formative.”

Admitting that such a decision is not easy at universities, the experts suggested that EPP directors can be appointed as vice rectors or advisors so that they are likely to obtain authority power in the management. In other words, some decisions can be taken in the university senate sometimes even without taking the views and opinions of the school director. Thus, it is noted that when the school director is in a close position with the rector, it may be more effective in preventing this ambiguity or providing accurate information and influencing the potential decisions. One of the experts explained how such a high-level status was significant in the following extract:

“As you may know, whatever the rector approves is realised in our country. Another point is whoever is close to the rector plays a key role in determining the requirements and

providing true-wrong guidance and information. The closer the school directors work to the rector, the more relevant and sound decisions they can take about the schools.”

Moving onto a more specific solution, the experts referred to the responsibility and power of testing units within the school. According to them, suggestions and comments of TUMs at these units should be taken on key issues about exams. One expert expressed her suggestion:

“The exam format is changed without our evaluation or view. From time to time, we are never consulted on whether the new format is suitable or not, and what are its shortcomings. We need to be present at the meetings, at which important decisions are taken, to express our opinions and to give approval on issues such as passing score, weighting, the inclusion of skills, especially speaking and writing skills, asking multiple-choice questions, and the announcement date of the scores.”

It is believed that the directors having worked in such units in their earlier careers may not disregard the consideration of TUMs. Bearing this in mind, experts stated that working in these units may lead to higher-level appointments by prioritising service in the testing units. They recommended that service in testing unit or another unit at EPPs may be required when appointing the school director or vice director. Therefore, the academicians who want to be appointed to a higher-level position at the school or university would be willing to work in these units. An expert, for instance, considered:

“Working in these units should be encouraged. If a financial income cannot be provided, it can be taken into consideration as a criterion in the career path. Considering that the kitchen work of an EPP consists of both giving courses and working in these units, an instructor who only gives lessons seems to have incomplete knowledge and skills.”

The fourth issue emerged as the necessity of coordination and orientation. First of all, coordination among all units at an EPP is a noticeable solution suggested by the experts. Apart from testing units, there are various units such as curriculum, material development, independent learning, learning counselling, professional development units, etc. within EPPs. Although they are formed in different structures, it is suggested that there should be good coordination among these units. As mentioned by an expert:

“Although they seem to have different responsibilities, they all serve a mutual mission: to carry out the education process in a healthy and effective way. These units may organize

annual, quarterly and monthly meetings. The meeting interval can be determined according to the needs of the school. In this way, there might be agreement on the decisions to be taken and the work to be done.”

Suggesting that all the units can work together to prepare an orientation programme about the importance of foreign language and can provide it to new students at the beginning of the year, the experts noted the importance of awareness of the newcomers. An expert elaborated on the necessity of orientation:

“It's sad to say our students still don't have enough awareness of the benefits of learning a foreign language. In summer, these units can work together to prepare an orientation for new students on the impact of language learning on their academic, professional and social life. In this way, instead of focusing only on exams, the students can be motivated more for courses.”

In addition to students, it seems that instructors need extra attention. As highlighted by the experts, it is suggested that the continuity of instructors' professional development in technology, field expertise knowledge, and testing may be fostered.

The last issue focuses on the provision of firm authority at EPPs. The relations between TUM-Instructors-Students should be maintained at the management level. The general views of the experts indicate that a lack of school management authority leads to the clash between TUMs and other instructors as well as students. Considering most of the TUMs complained about their colleagues' inappropriate attitude during the examination process, even from the pre-testing to post-testing, one of the experts stated:

“That students or other instructors come to the testing office and make complaints easily can cause both possible undesired events and concentration problems even in its most natural state. Therefore, the school administration should have an action plan to establish an official process for complaints or feedback.”

Accordingly, TUs may be authorised for complaints/objections unless there is a breach of duty. The experts focused on the unbearable workload of TUMs due to unjustified objections of students. Therefore, they suggested that the decisions should be taken from TUMs even without asking for permission from the school director. An expert elaborated on the issue:

“When students fail in the exam, they tend to complain to the administration in order to get higher scores. We need to find their papers and start the evaluation process from the

beginning. We do not want to take their rights from them, but the objections should be based on solid justifications. Since we are experts in testing, we should be authorized in return.”

Another suggestion was that sanction decisions must be taken on breach of duty, inadequate proctoring. Inappropriate behaviours of the instructors were reported to result in other unexpected problems during the implementation of the exams. As explained by an expert:

“Forbidden attitudes and behaviours during the examination such as playing on the phone, sleeping, reading a book, doing homework, chatting with a friend or leaving the classroom must be sanctioned, otherwise the authority gap gradually increases.”

The findings of the study reveal that there are 20 codes as solutions in the second category, namely institutional solutions. The experts noted the significance of training, physical and technical needs, structural priorities, coordination and orientation, and firm authority. The first four codes included digitalisation of exams, pre-service and in-service training, priority to TUMs in assignments, and collaboration with national and international organisations. Besides, it was suggested that TUs should be supported in terms of physical conditions, technical staff, licenced programmes, and compatible automation system. Key solutions were provided on division of education period, system employed, appointments of EPP directors in key positions, importance of services in similar offices, and participation of TUMs in key decisions. In addition, they proposed coordination, orientation, professional development, and a common calendar. Furthermore, maintenance of relations of TUMs-instructors-students, giving authority to TUMs on complaints, and sanction decisions were suggested, which could strengthen their power.

4.2.3. Unit-level solutions

Unit-level solutions comprise the least frequently stated type of solutions (N=6.197; 32,02%) to the challenges encountered in testing processes at EPPs in higher education. This category includes twenty-five emerging codes which are shown in Table 4.23.

Table 4.23. *Unit-level solutions*

Emerging Codes	Number of Codes	%
Transferring 21 st -century skills into exams	21	9,95
Using electronic open sources	20	9,48
Separating philology students	19	9,00
Setting up a bottom-up circulation system	18	8,53
Preparing a table of specifications	17	8,06
Working on validity and reliability matters	17	8,06
Writing well-organized instructions	16	7,58
Choosing up-to-date and authentic exam topics	16	7,58
Determining task and responsibility sharing	15	7,11
Preparing a clear workflow	15	7,11
Setting up a dynamic proofreading cycle	14	6,64
Giving orientation to newcomers	13	6,16
Receiving training from online platforms	12	5,69
Having membership in related associations	12	5,69
Working on standardisation	11	5,21
Determining reliability processes	10	4,74
Preparing and updating well-developed rubrics	10	4,74
Working on risk management	9	4,27
Collaborating with TUs at other universities on exam security	8	3,79
Leaving a free room for meetings/group work	8	3,79
Receiving online feedback	7	3,32
Employing statistical analysis	6	2,84
Working on benchmarking	5	2,37
Developing physical and online archive system	4	1,90
Preparing an item pool	3	1,42
Total	211	100

The first solution suggestion is related to systemic system comprising the most frequently used type in this category in the interview. The experts seemed to believe that the exams should be prepared regarding 21st-century skills. An expert elaborated on his solution as follows:

“We must ensure that our students become aware of their communication skills by transferring communication skills and other 21st-century skills into our exams.”

Looking at the general picture, the experts stated that the resources available in testing units were limited and the worse thing was that the existing ones were insufficient and old resources to transfer 21st-century skills. Thus, it is recommended that TUMs be encouraged to use electronic open sources, considering how difficult it may be to supply new resources due to economic and other reasons. One of the experts specified her suggestion as follows:

“We find very good open resources for texts, reading, listening parts but we have not been able to convince the director. This is the case at many universities. He believes students can find these sites, read texts beforehand, which might be against test security. However, open sources are both economical and useful.”

As shared by the experts, it seems that the use of electronic resources may contribute to digitalisation process within these units. Assuming that a number of students cannot come from different cities earlier for the placement exam, which is conducted earlier than the education period, the experts suggest placement exam be delivered online so that the students do not have to look for temporary accommodation until the education period. The following extract shows an expert's suggestion:

“The term starts 1-2 weeks later than the placement exam. The students who come to the exam must either find a place to stay during this time or go back to their hometown. Therefore, there are so many students who don't take the placement exam. Whereas, if we do it online, the number of those who take the exam will increase and those who cheat will decrease over time.”

Choosing the sources for different student groups may put extra challenges. Therefore, most of the experts suggest that philology students are suggested to be separated from the general group. One expert expressed his solution:

“In terms of the balance in the classroom, the level difference between philology students and others is very high. This difference harms in-class level harmony and creates different profiles. Therefore, we should separate at least these two groups from the exemption exam.”

Differentiating the groups might increase the workload of instructors. However, setting up a bottom-up circulation system in 5-10 years is believed to ease the burden. As exemplified by an expert:

“Dealing with the unpaid workload without financial or non-financial rewards leads TUMs to the burn-out syndrome. Therefore, the testing team should be changed after a certain period of time considering the number of personnel at the school.”

The second issue accounts for language testing issues. One of the most emphasized issues in the focus group discussions was to carry out some studies before the exam preparation. First, the experts indicated in the discussions that a table of specifications suitable for needs of the relevant institution and students should be prepared. For the table of specs, it is recommended to create a roadmap, including learning outcomes, units, and subjects through meticulous joint studies and meetings with the curriculum unit, level coordinators, and other relevant units. One expert expressed his solution:

“A Table of Specifications aims to identify areas of success and ensure that the exam is created fairly. It provides the construction of a test focusing on key areas, the weighting of learning outcomes appropriately based on their importance, and content validity of a test along with evidence of what needs to be covered.”

Second, another point made by the experts was about validity and reliability matters. It seems necessary to make a plan about which types to what extent face, construct, content validity kinds along with internal consistency and interrater reliability will be taken into consideration in order to ensure that tests are sound and replicable while the results are accurate. An expert cited her suggestion:

“Reliability and validity are the two fundamental features to assure the integrity and quality of a measurement instrument. They not only increase transparency but also decrease tester bias in exams. Lack of assessing reliability and validity, it may be difficult to identify the impacts of measurement errors.”

Third, in addition to the solutions stated above, writing appropriate instructions is a key factor contributing to the smooth testing process. This issue is two-fold: students and instructors as proctors. According to the experts, clear instructions are likely to ensure that the students fully comprehend what they need to do to achieve in the exam, which can ease their nerves, assuage their insecurities, and help them confirm the expectations. On the other hand, the experts stressed that clear instructions for proctors are advised so that the instructions would give them a good idea of what is expected. As exemplified by an expert:

“The instructions are expected to have a clear and precise language in which short, complete sentences with precise, concrete terms are used so that students are able to grasp what's expected of them. For instance, assuming that an assignment is large or multifaceted, it can be sectioned into smaller tasks, in other words, manageable tasks, for students who can't handle a long list of directions, which may help students feel more confident. Besides, clear instructions should be written for proctors, regarding their punctuality for the assigned tasks, their disengagement in any other activity other than monitoring the examination and their manner of response to students' problems and raised hands without any verbal or nonverbal hints regarding the correct answer.”

Fourth, the experts explained that sometimes students seem to have problems perceiving and understanding the subject in the exam, no matter how clearly the instructions are written. It is believed by the experts that factors such as the mismatch between the selected topics in the exam and students' interests and attention as they may be outdated play a major role in the failure of the students. Thus, it is recommended that exam topics be chosen from up-to-date, authentic and appealing issues to students. Fifth, the experts indicated in the focus group that task and responsibility sharing should be determined before the term; otherwise, it would be more challenging to set a system during the action and rush, which might cause ambiguity in the process. Sixth and similarly, they noted that a clear workflow must be prepared to show the timeline of the tasks and responsibilities in each task. The last issue to be regarded before the test preparation phase dwells on proofreading. As stated by the experts, a dynamic and easy-to-follow proofreading cycle is a significant improvement allowing the TUMs to edit the tests and avoid individual errors and mistakes. To the experts, proofreading seems to ensure that the test items are clear for delivering the planned activity in the interest of the target group through reading content with caution and marking errors if exist. An expert explained how she perceived the necessity of proofreading elaborating on:

“To put it simply, proofreading provides an overview of the already written test with an opportunity to review the work by re-reading with a fresh perspective what test writers have generated. It ensures that the test flows well, doesn't cause the students to stumble, and conveys a definite meaning. For a simple example, the next reader focuses on surface errors such as misspellings and mistakes in grammar and punctuation.”

The third issue is about specialisation in language testing. It is believed that specialisation can be initiated through the delivery of orientation from the experienced TUMs. Accordingly, all the experts stressed that orientation should be given to newcomers to the TU. An expert, for instance, considered:

“A guide can be prepared for general information on testing. Besides, an orientation programme that explains testing procedures and principles of the university can also be useful for newcomers in order not to be shocked.”

The experts also noted that the opportunities at a university could be limited; however, this should have prevented them to receive training from online certificate platforms. An expert explained as follows:

“With the outbreak of COVID-19, the number of online education opportunities has increased rapidly. In fact, it is wrong to expect everything from the institution. There are many paid or free online courses on testing so that TUMs can develop themselves.”

In addition, the experts added that there are several international or national associations engaged in language testing. They recommended that TUMs could be registered with those associations such as FOAI, Metasig, and ALTE. As mentioned by an expert:

“Membership in national or international associations enables TUMs to receive information about activities on different subjects on testing. In this way, they can both meet their peers who are members of the association and benefit from their experiences. They can also discuss their problematic issues at organized events and find a common solution.”

The fourth issue is about evaluation issues. One of the major challenges during the evaluation was believed to be standardisation during the interviews. According to the experts, TUMs need to work on standardisation so that a consistent evaluation system could be adopted by raters. The following extract exemplifies the issue:

“When standardisation meetings are avoided, personal differences in scoring increase, which is supposed to risk the reliability of the examination. Therefore, open-ended questions are avoided in exams due to concerns in assessment. In order to ensure reliability, standardisation meetings and orientation should be provided to all raters”

Accordingly, the standardisation problem seems to result in discrepancies in the assessment. Therefore, the experts stressed the importance of reliability processes which should be determined beforehand. An expert cited her suggestion:

“Considering that the interpretation of the papers differs according to instructors' point of views in the absence of reliability measures, TUMs should establish a well-planned system for reliability.”

To ensure an appropriate reliability system, it was advised that well-developed rubrics needed to be prepared and updated to fit the needs of the institution and students. One of the reasons underlying discrepancies mentioned above was linked to a lack of suitable rubrics. The following extract would be enough to clarify the issue:

“TUMs should analyse different types of rubrics, nationally or internationally used ones in order to develop their own rubric. They can also make adaptations to the existing ones to suit their needs. In addition, the rubric needs to be revised after a certain time. Besides, all the raters should be informed of the rubric designed; otherwise, they may have difficulty in interpreting it correctly.”

The fifth issue is about test security. The experts recommended that TUMs had to work on risk management. They were suggested to benefit from the criteria displayed on the webpage of quality assurance agencies. One of the experts stated that:

“Regarding system failure, TUMs can have a backup of the exam papers or scores. One suggestion could be to scan the papers scored and save them on an external hard disk. They could also send exams to their e-mails or store them in Google drive.”

It was considered that they might not know what to do with risk management on their own due to a lack of experience and training. Then some experts offered to collaborate with TUs at other universities on exam security. The following extract exemplifies the issue:

“TUs can form a WhatsApp group into which all volunteer testing members could be added. In this group, they may ask each other about challenging issues including exam security.”

It was concluded in the focus group that exam security could not be maintained without preventing sudden visitors i.e., students and their colleagues to the office. Therefore, they proposed that a free room should be given to the testing office for

meetings/group work. This was considered to contribute to not only exam security but also their concentration during exam preparation. An expert explained as follows:

“Never-ending traffic in the office might risk exam security. Assuming that exam papers would stay on the table before the exam, and a visiting student might see the papers and take one sample, who will be blamed? Therefore, it is of utmost importance that a free room should be left for meetings and visitors.”

The last issue is about the post-evaluation process. Some experts firmly asserted that language teaching, learning, and testing were turning into a cycle in which one affected the other. Therefore, it was suggested that a feedback mechanism should be established to receive online feedback from students and instructors following the examination. An expert elaborated on his solution as follows:

“Sometimes feedback is considered a criticism process. However, it is necessary to identify the weak points and discrepancies of the system. Thus, a constructive feedback cycle should be formed to develop the efficiency of both language teaching and testing.”

Experts mentioned that another feedback source could be a statistical analysis of exam results, displaying descriptive information including average scores. It was believed that this analysis could help match the results between the groups, classes, and students in the same class. Besides, they added that there might be a comparison of the scores of the current year with those of the previous years. Thus, they highly suggested employing statistical analysis on exam results and items. One of the experts specified her suggestion as follows:

“EPPs should be delivered staff positions for experts equipped with technical knowledge for statistical analysis. In addition, in-service or pre-service training should be provided to other TUMs on statistical analysis. Furthermore, licences should be taken for the software programmes in need.”

It was a consensus among the experts that work on benchmarking with national and international exams could contribute to the testing process of EPPs and the quality of exams. Benchmarking was thought to be improving performance through efficiency by cutting out waste in the processes, decreasing costs, time or effort spent. They also believed that it would ensure streamlining the operations and ultimately help TUMs retain more of their revenue. One of the experts touched on the benefits of benchmarking:

“Through benchmarking, TUMs can gain an independent perspective about how well they perform compared to other tests so that they may drill down into performance gaps to identify areas for improvement. This ensures the development of a standardized set of processes and metrics. Besides, it enables a mindset and culture of continuous improvement in TUs, which helps TUMs could set performance expectations.”

The analysis of the focus group meeting revealed that the key reason to archive documents was to avoid losing data since documents were vulnerable to being destroyed or corrupted, either by accident or by such a natural disaster as a flood or fire while electronic documents could be compromised by security threats. Therefore, they proposed to develop both physical and online archive systems. As exemplified by an expert:

“By establishing a secure archiving system of documents, TUs can increase protection from unauthorised third parties and reduce the risk of files missing. Archived documents provide evidence, explanation, and justification not only for past but also current implementation.”

Last but not least, the experts centred their focus on the importance of item pool for more quality exams. They stated that item pools would improve the chance that exam takers do not see the same set of test items for every test administration, preventing cheating in the system. In this sense, they proposed that TUMs would prepare an effectively and efficiently working item pool. One of the experts elaborated on the solution:

“A well-constructed item pool is an essential part of the examination process. One significant point of an item pool is that enough items need covering the scope of the content domain. They should get harder or easier depending on how a student is performing on the test. Therefore, it must adapt to measure on-, above-, and below-grade abilities so that instructors are informed about each student’s readiness.”

The analysis of the study showed that 25 codes emerged as solutions in the third category, namely unit-level solutions. The experts touched on systemic changes, language testing issues, specialisation, evaluation issues, test security, and post-evaluation. The first four codes included 21st-century skills, open resources, separation of philology students, and bottom-up circulation. Besides, it was suggested that TUMs should work on table of specification, validity, instructions, exam topics, task sharing,

workflow chart, and proofreading cycle. Key solutions were provided on orientation to newcomers to TUs, training from online certificate programmes, and membership to assessment organisations. In addition, they proposed that TUMs should ensure standardisation, reliability, and well-developed rubric. Furthermore, work on risk management, exam security and provision of a free space for meetings were suggested. Last, they were offered to employ online feedback, statistical analysis following the exams, benchmarking with closest universities, online archiving system, and item pool.

To sum up, seven field experts participated in the second phase of the study to identify common and feasible solutions to the challenges determined in the first phase. The findings of the second phase show that the experts proposed three main categories as solutions, namely national, institutional and unit-level solutions. The first category was about national-level solutions, including 15 codes. Major codes were related to legal studies, comparable skill-based exam, financial and non-financial incentives, and bottom-up needs analysis. The second category consists of institutional-level solutions, with 20 codes. Major codes were related to training, physical and technical needs, structural priorities, coordination, orientation, and firm authority. The last category includes the solutions at the unit level, with 25 codes. Major codes were related to systemic changes, language testing issues, specialisation, evaluation issues, test security, and post-evaluation issues. Figure 4.7. illustrates key concepts drawn from the findings of both phases.



Figure 4.7. *Key concepts*

Taking into the challenges and solutions drawn from the two phases, the researcher concludes that eight key concepts emerge. The key concepts consist of legal ground, washback effect, standardisation, training/specialisation, emotional factors, including motivation, burnout, and anxiety, technical, physical, and human infrastructure, sustainability and efficiency, and digitalisation.

As Figure 4.7. shows, the key concepts the researcher designs seem to be interconnected and have impact on one another. All these key concepts will be discussed in the following section.

CHAPTER 5

5. DISCUSSION

This section draws upon a discussion and an interpretation of the research findings derived from the research questions and interviews along with the related literature, puts forward implications to various stakeholders, namely authority bodies, university/school leaders, TUMs, and concludes with suggestions for future research.

The study was dependent on two research questions in that the discussion terms of the second research question were shaped by the findings of the former one. Within the scope of the first research question, this study investigated the challenges encountered in the language testing processes of EPPs in Turkish higher education institutions. Seeking an answer to the second research question, it also attempted to search for feasible solutions to the challenges. In order to reach these objectives, the study is designed to include two phases, namely one for the challenges and the other for solutions. The findings for these research questions were obtained through the administration of semi-structured interview questions to 208 participants (73 directors and 135 TUMs) in the first phase and focus group interview with field experts. The obtained data were analysed descriptively and inferentially. The systematic analysis of the emerging codes both in the interviews with directors and TUMs and in the focus group with experts has two main contributions. Data analysis unpacks the challenges encountered in TUs and their impact on TUMs' work in the offices, illustrating their struggle and unreturned sacrifice. Findings also indicate that suggestions for improvement cannot be achieved by TUMs or school directors alone, but there is a need for multi-layered change. This means that no matter what TUMs change in their office, it would not lead to sustainable and systematic solutions. Thus, relevant policy needs to be reframed, and its scope broadened, to better comprehend the long-term consequences for not only TUMs and school directors.

In the first phase, interviews were conducted with directors/vice directors and testing unit members separately in order to find out the challenges in testing processes through a semi-structured interview form. Including school directors in the interviews helped to integrate the managerial challenges into the data collected. Likewise, the diversity both within and across the two groups of research participants contributed in turn to exploring both the similarities and differences within and across each group, thereby leading to bringing together multiple points of view of a multiplicity of

participants in the findings of the study, in the exploration of the challenges in testing processes and processes at EPPs of higher education institutions in Türkiye. In this part, the participants (N= 208) were selected from diverse universities determined according to some certain parameters such as development, legal status, establishment year, special mission, and so on. Employing the strategy of diversity in the selection of universities (35 state universities and 15 non-profit foundation ones) served in the data collected to cover different challenges some of which may be peculiar to a specific group of universities, increasing the strength of the sample and the usefulness of the findings. Furthermore, conducting interviews with two former directors helped to compare the difference between one in an active position and the other having left it and incorporate the role of active working into the data. The systematic deciphering of the emerging codes in the analysis of the data through a reconstructive process helped to place the research participants as an integral part of the analysis, that is, helped to centre the focus on the mutual challenges in language testing processes at EPPs.

In the second phase, solutions were sought with seven field experts through the Delphi method. Selecting experts for this method was crucial since they should both have enough experience in language testing having worked in these offices beforehand and be open to discussing their opinions without any bias and pursuing any hierarchical superiority or advantage over other experts. Besides, employing this method in the second phase served to process the suggestions in three steps, giving the experts a chance to get familiar with the challenges found in the previous phase in the first part of the method, initiating brainstorming and building a map of ideas through discussions on the key words by the experts in the second part of the method, and finally asking them to reconsider the common solutions, allowing them to revise the existing ones and add if any new idea springs to their mind in the last part of the method.

Based on the voices of the participants, findings seem to indicate that the challenges are connected with each other influencing the impact of the other and recurring the challenge for a long time, which in turn seems to worsen the working conditions of TUMs in TUs. The results of this study indicated that nearly all TUMs had negative perceptions about working in TUs and improving language assessment processes in their schools. The researcher realised that a number of TUMs felt relieved while reflecting on the dead-end challenges and learning the similar ones emerging in other universities. Besides, it was found that the interviews yielded enthusiasm among

directors and TUMs about applying suggestions of this study for future development. Similarly, some TUMs were curious to realise the importance of language testing and assessment as a scientific study, thus mentioning their willingness to study it in their MA or doctorate studies. Therefore, the study puts forward that as studies on the challenges within TUs increase, this issue will attract the interest of more people and eventually the attention of those who work in authority institutions.

Taking into account the perceptions of the participants in the first phase and suggestions and comments of the experts in the second phase, the researcher attempts to discuss the findings around eight key concepts as follows: legal ground, washback effect, standardisation, training/specialisation, emotional factors (motivation/burn-out/anxiety), technical/physical/human infrastructure, sustainability/efficiency, and digitalisation. What follows thus presents a discussion of the challenges by portraying their importance within an organisation, presenting the participants' disadvantages, and highlighting experts' suggestions for improvements with regard to the challenges. The section concludes with a reflection on the challenges and solutions that took place after the completion of the data collection.

From this perspective, the core challenge borne out in the data seems to be a depiction of legislation requirements by the participants. Although this challenge is not directly involved in the practices within TUs, it is thought that the need for a legislative change at the national level lies at the root of other lower-level challenges. Although each phase of the study had a different aim and methodology, the results of each phase are discussed together in order to ensure that they both support each other.

5.1. Key Finding 1: Legal Ground

The first concept to be discussed is the utmost necessity of legal ground. During the interviews, all of the testing members reported that the unit where they were working was not recognized officially and had no place in legal regulations, which had a negative impact on them. In other words, there is no such a unit on the legal grounds. They firmly asserted that if legal regulations were established regarding the testing unit, at least mandatory minimum limits could be determined. Maybe much workload would not be given or they would get reward(s) in return. Yet, currently, they are not paid back

since their duty has no legal basis and is perceived as an extra burden. Bourdeaux and Chikoto (2008) claim that legislation plays a crucial role to carry out reforms in institution management. On the other hand, when the necessary legal ground lacks, nearly all other challenges seem to remain still. This situation seems to make the participants more reluctant to work in the office, which is believed to have the potential to block the whole system of EPPs. This reluctance appears to be connected with their poor job satisfaction. Kocman and Weber (2018) describe job satisfaction as an attitude toward one's job consisting of evaluative judgments about that job, including affective-cognitive work experiences and beliefs about the job stemming from a subjective evaluation of the relevant job. In this case, it can be thought that instructors' concerns due to legal loopholes seem to severely influence their feelings and views about possible duties in TUs as a result of their subjective evaluations. Widarko and Anwarodin (2022) also approve of the importance of work motivation in their performance and their reluctance or willingness.

Similarly, there is no such title as a testing unit member though the ones working in TUs are called testing unit members by the management, their colleagues, and even some students. Due to the legal loophole, testing members' job description is not clear as there is no post as a testing unit member in official terms. Timothy and Ryan (2008) associate one's work with their identity. They explain their association illustrating the moment when people are asked a general question as to what they do, they generally respond with their job title. For example, there is a job description of the manager, the workload is given according to the job description. Not only financially does he get a promotion after that position because of having a title. However, TUMs' office work does not take them anywhere. This clearly shows that job description occupies a central role in their attitude towards their work. The result of this situation might be examined from two perspectives. In the case of TUMs, they may appear to be unwilling to work in these offices and prepare quality exams. Accordingly, job satisfaction may be related to workplace beliefs and behaviours, affecting even life satisfaction. The attention of scholars (Harrison, Newman & Roth, 2006; Lee, Park & Koo, 2015; Ng, Yam & Aguinis, 2019; Woznyj et al., 2022) has centred on the relationship between general job satisfaction and some significant factors such as job attitude, organisational commitment and individual effectiveness including workplace behaviours, performance, lateness, absenteeism, and turnover. On the other hand, in the case of the directors, they

may appear as bad character in the movie, forcing some instructors, especially the novice ones to accept the assignment. In a study conducted by Gümüş and Gülmez (2020), it is stated that middle-level managers, defined as academic leaders, should develop good relationships with their senior management bodies, instructors, and experts. However, they also stated that in some cases their powers and action possibilities were limited.

At this point, it is crucial, as experts suggest, that the legal status of not only the testing unit but also other units such as curriculum, material development, etc. be determined in the relevant legislation. While doing so, the titles of a testing unit member, material development member should be cited in the law along with their responsibilities and duties. What appears to be important for them is their rights when compared to the instructors who are employed for teaching. In other words, their rights should be equal to those of teaching instructors, which might prevent any discrimination between the groups. Demir, Kapukaya and Özfıdan (2015) found that instructors working at EPPs encounter problems in terms of working conditions and social rights while doing their jobs. In addition, extra difficulties are experienced when they are required to work in an office such as a testing unit. Taking all into consideration, improper career options as well as unfitting labour conditions are highly correlated with reduced job satisfaction in the general population and can lead to reduced quality of work. Accordingly, Aktaş and Gündoğdu (2020) stress that instructors' roles and responsibilities are explicitly defined to run EPPs more efficiently. By the way, it was found that in several universities, instructors both taught courses and were requested to work in offices such as testing units at the same time. Therefore, it was recommended by the experts that the course load of those who would be working in offices in addition to teaching should be determined in the law as less than solely teaching instructors.

From this perspective, the issue seems to have been approached from a narrow viewpoint without tackling the bigger picture of the power of EPPs within the university. According to the responses of both directors and testing members, it was noted that EPPs are not or barely involved in the decision-making process in the university, which weakens the representation power of EPPs neglected in universities. It is believed that due to the weak voice of our school in the Senate, TUMs experience problems and have to solve them on their own, but they end up feeling in a desperate situation. Decision-making plays a crucial role in the functioning of an organisation

(Kumar & Gautam, 2018). However, one director harshly criticised his situation, comparing EPPs and Continuing Education Centre (SEM) and noted that EPPs were less important than any unit, even SEMs, when it came to serious decisions and projections related to the university or the school itself. He complained about the ignorance of the school in determining even the academic calendar, which leads to a number of unexpected students, exams, incidents, and so on. Decision-making is believed to show how a manager works on challenges and tries to find a solution to resolve them (Aboudahr & Olowoselu, 2018). Therefore, school directors are expected to cope with unexpected challenges although they have limited power. It was like irony when one participant mentioned that university leaders suddenly took the decision of moving the school to the north campus and informed them on Friday in September when they were very busy. They were only told to pack their things during the day. In addition, another TUM complained about a reduction in the weight of the final without being informed beforehand since the Vice Rector of the university wanted to decrease the number of students, but there was no one from the EPP in the committee when the decision was taken. The sad part dwells on the fact that most of the representatives were unaware of the internal processes of the EPP. However, that committee decided on its own, without asking anyone from EPP or TU, even the director. The approval was already given by the administration.

The experts seem to voice the assignment of the director of EPP as one of the vice rectors or advisors, which to me would contribute to their relevance to the rector and decision taking meetings. According to Turkish higher education legislation, a rector has a major role in a university (Doğan, 2018); therefore, the close relationship between EPP director and the rector would ease the submission of real knowledge and demands. In addition, it was also suggested that the director of EPP was assigned amongst the candidates who have served in one of these units beforehand. In organisational culture, an organisation can be successful through quality decisions as effective decisions can be taken by qualified managers (Iqbal, Akhtar & Saleem, 2020). The crucial point is that the representative impact of EPPs should be increased within the high-level authority, including the Senate.

The limited participation of directors of EPPs in decision-making seems to account for why EPPs are provided limited financial contribution or support although they need sources, technical equipment, or software systems. About two-thirds of the

participants in the sample reported that there was not much financial support for testing although there was a need at a lower level, namely in the kitchen of the school. It was indicated that if there was a demand from the bottom, it was obvious that there was a need. Although they listed the must-buys, they were told that the resources of the university were insufficient. Therefore, they are compelled to work in these offices by persuasion, by force, or by their own intrinsic motivation, though very rare. However, it is interestingly sad that universities spend a huge amount of money on irrelevant and nonessential organisations, which is firmly believed to violate the public interest. For Küçükoğlu and Taş (2022), continuous increase in unnecessary expenses will lead to long-term economic problems that eventually affect the whole society in a negative way. It can be alleged that due to unnecessary expenses in universities, EPPs cannot get crucial support. Hence, the experts recommended that incentives could be provided as a reward in some way to the instructors who work in similar offices in addition to their teaching duty.

Another unique factor contributing to the disrepute of the testing unit among instructors is the load of uncompensated extra work. Beckers et al. (2008) claimed that long-term unpaid and forced tasks cause health problems in employees such as stress, hypertension, diabetes, heart disease, and sleep disturbance while negatively influencing their well-being. The results of this study show that testing members appear to sacrifice their energy and time, even their free time and off days, in order to complete exams on time and keep the intense flow of the workload. Timothy and Ryan (2008) assert that there is a strong and consistent relationship between one's job satisfaction and subjective well-being. Even ignoring their well-being, testing members were observed to have mental fatigue. According to Techera et al. (2016), mental fatigue leads to a loss of motivation in current activities and weariness. Therefore, testing members' mental fatigue is supposed to decrease their ability, competency, and productivity to process their tasks, diminishing their error avoidance. What seems to be paradoxical is that they are harshly criticised for a minor error though they leave aside their graduate studies or steal time from their families for their work. In addition, they are given an unofficial duty and responsibility, but not authority. This is supposed to leave them desperate against the clash with other instructors, students, or managers from time to time. Taking these points together, it would be argued that such extra work without financial compensation causes the units to remain understaffed. It also appears to provide an

explanation for the reason why instructors try to avoid being assigned unless they are a novice instructor. A testing member drew the attention to the unfair situation between instructors and testing members and stated that the instructors would teach their courses and leave as soon as possible while she was struggling to take printouts and shuffle the papers for over 2000 students until midnight. Huang et al. (2019) associate job stress with an employees' interaction with their psychosocial environment at work.

Building on experts' comments, it is suggested that a well-planned incentive system be established by authority councils. Thibault-Landry et al. (2017) suggested that financial incentives contribute to the feeling more competent and autonomous, fostering greater autonomous motivation and better work performance. Academic incentive initiative led by Council of Higher Education might be revised in that the service in units could be awarded some points. Another possible incentive may be new specific project calls by the Scientific and Technological Research Council special to language testing and the ones working in these units. A third initiative might come from Student Selection and Placement Centre that could include testing members working at universities in its item writer pool. For Bryant and Allen (2013), not only financial incentives but also non-financial incentives as a combination of rewards are more effective increasing work attraction and motivation of employees within an organisation. Therefore, testing members can be awarded certain extra points in accordance with their service year when they apply to Inter-universities Board for the associate degree. In addition to national incentives, at the institutional level, universities would give priority to the academic staff working at units in the assignments for international and national events and duties. As a non-financial reward, employees can be provided priority in training and personal development (Schlechter, Thompson & Bussin, 2015). The easiest one might be Erasmus Staff Exchange through which testing members can visit similar units or centres in European universities. It is also suggested that the exams, not small-scale, prepared by TUMs be displayed on their official academic page, that is, in their YOKSIS account, which is believed to contribute to the visibility of their work. Last but not least initiative may be the freedom to work from home office in a certain amount. With the use of information technologies in working life and the emergence of the COVID-19 epidemic, which broke out at the end of 2019 and spread to all countries all over the world in 2020, the remote/home working model

was adopted by institutions (Mustajab et al., 2020; Raišienė et al., 2020; Savić, 2020; Serinikli, 2021).

One other challenge is about the responsibilities of instructors who fail to meet assessing deadlines. In addition, most of the participants argued that instructors were reluctant to get exam proctoring. Although a proctor is absolutely necessary for each exam, proctoring duty is undesirable in universities (İlkuçar, 2011). They were claimed to miss packs, tend to take medical reports just before the exams and hand in papers late. Most of the TUMs were of the opinion that their inadequate proctoring stemmed from a lack of punishment or sanction. This is in line with the suggestion of the experts. According to them, an official sanction system as punishment might be legalized for proctoring and other responsibilities regarding the examination. I believe even such an official attempt would deter instructors from breach of duty. Experts also highlighted that authorisation should be transferred to the head of the testing unit in the case of objections and complaints from students or instructors unless there was wilful neglect, which is supposed to prevent unjustified objections and complaints.

Considering all the challenges related to the lack of legal basis mentioned above, TUs are understaffed in almost every university. Some TUMs complained that there were more than 1000 students per TUM. Others added that only one person was responsible for one level. Therefore, it was noted they could not have off-day even when they had health or family problems. At this point, it is important to make office tasks in EPPs attractive. According to the experts, providing both financial and non-financial incentives by law to instructors in these schools will be a source of motivation, which is believed to increase the number of volunteers to work in TUs and similar units. Ogunode and Okwelogu (2022) listed underfunding, poor motivation, and an uncondusive working environment as responsible for the shortage of necessary academic staff in universities.

All in all, the interviews revealed that TUMs in many universities preferred employing multiple-choice questions instead of open-ended questions due to understaffing and the washback effect of the national exams administered by ÖSYM. This dead-end effect is explained in detail in the following subsection.

5.2. Key Finding 2: Washback Effect

The second concept centres around the washback effect of high-stake exams, namely YDS, YÖKDİL whose names may change over time (previously KPDS, ÜDS) but not their content and main spirit. One interesting observation about the findings concerning this effect is that the exams prepared at EPPs appear to have transformed into test-based exams and the schools are considered to be working as private language courses that prepare students solely for tests though their initial and their core intentions were to equip the students with the necessary language skills and abilities by paying more attention to mechanical skills. Salehi and Yunus (2012) stressed that university exams are negatively affected by the washback effect of test-based exams. A testing member approached the issue by emphasizing the similarity between their exams and those prepared by ÖSYM and noted they felt as if they were working at the centre. Despite any efforts to make tests well-planned, the skills required to become successful on a test may not fully equate to the ones needed to succeed in a target language use (Bachman & Palmer, 2022; Bachmann & Palmer, 2010). It is fair to conclude that in this way, the purpose of language education is deviating while the focus changes to passing the exam(s) by memorizing mechanical information rather than learning the language and acquiring skill-based abilities. This deviation causes students who are used to the test system to focus on passing the exam(s), instead of learning the language in general. In such a test-based system, the limitations including time and space force test developers to be selective in what they are testing and at the same time to be pragmatic in how they are conducting the test (Green, 2013). In this sense, the experts argued that the decision on which exams would be equivalent should be left to universities, which might vary from one university to another depending on their students' backgrounds, needs, and aims of the institutions. They also reported that test-based exam scores could be avoided with a senate decision. Considering the consensus among the experts, it can be deduced that such flexibility on equivalence decisions would contribute to the autonomy of universities which as defined in methodology differ from each other in terms of their region, characteristics, students and staff numbers, etc. Besides, the experts strongly supported the transformation of test-based exams such as YDS, YÖKDİL into skills-based exams including four skills. This transformation may be painful at first, but considering that there are sufficient human resources, namely foreign language instructors working at Turkish universities,

considerable success can be achieved in a short time. TUMs in universities might play a great role in exam preparation while the experience of ÖSYM in exam administration. In addition, a remarkable point touched on in the interviews was that there were a number of instructors who have received expertise certificates in language testing from international organisations and have taken part in international exams. Thanks to such advantages, deficiencies, and mistakes experienced in the preparation and implementation of four-skills exams can be fixed so that better quality exams can be formed in a short time. After a while, it can be considered that international accreditation is obtained. Then the exams prepared by a Turkish organisation can be used in different countries, especially in Africa, Asia, and the Middle East, which provides economic profit to Türkiye.

As another force of washback, most of the testing members in the interviews complained about the number of exams from quiz to proficiency. Therefore, due to a great number of tests throughout the year, it is rather challenging for them to be selective and pragmatic. Besides, students make their preparations including their self-studies in accordance with the exams they are provided when the assessment is based mainly on grammar or a test-based system (Paker, 2014). In the interviews, it was highlighted that even the number of proficiencies was increased to allow students to try their chance of guessing the correct alternative. Considering that there are three major criteria for quality in assessment and testing processes: validity, reliability, and usefulness (Brown, 1996; Ertürk, 2015; Rashidi & Javanmardi, 2011), it is expected that language testing would show how well the process works, is appropriate for its purpose measuring what is desired to be tested, and when it is repeated, it should give similar results in future situations. As mentioned above, of the seven experts, all actually suggested that a four-skill-based exam at a national level be prepared and implemented by ÖSYM in order to reverse the washback illness. In addition, it is fair to propose that equivalence and acceptance of the high-stake exams could be delegated to the decisions of the universities. A director openly shared the name of the university, which does not accept the scores received from the centralised exams, but he was also curious how possible that university does not but they do.

It could be claimed that this washback situation seems to have a domino impact for all reasons of the failure of language education in Türkiye. The responses of the participants signal that the quality of K12 education seems to decrease year by year,

claiming that the students having passed from K12 to EPPs show less success in proficiency when compared to previous years. One director at a private university shared some general results regarding the success rate of the students in the proficiency exam held in September. Of the 1400 students who took the fall proficiency at the beginning of the term, only 150 students (10,7%) received a passing score. What is more striking is 76% of the remaining started from the A1 level. Similar figures were shared by a director at a state university. She explained that they had too many students who took their fall proficiency, namely around 3100 students. Of those students, 450 students (14,5%) received a passing score while around 2100 students began fall education from A1-A2. At this point, the government, as experts highlighted, should consider ensuring the cooperative study between MEB and YÖK so as to deal with the issue through a bottom-up project starting from primary education to graduate programmes. It is believed that deterioration in K12 education leads to a decrease in the quality of incoming student profiles. It seems that the low level of students coming from K12 results in decreasing the difficulty level of the exams, avoiding the variety of subjects. According to the comments of the experts, it seems evident what language tests measure needs to be investigated from multiple perspectives. Focusing on K12 students' attitudes might provide further insights to decision-makers in selecting the most appropriate type for their aims, interests, and needs in given settings.

One major issue is related with the overdose of the test-based system in EPPs instead of learning-based. Most of the TUMs claimed about the number of exams due to demand from the management, which appears to shift the evaluation of the exams focusing on how many students pass or fail as they generally mentioned not having any other evaluation with less attention to language acquisition. Every instructor might have heard that students decided to study only when there was a quiz or exam. It was also frequently complained that due to understaffing, the existence of high-stake exams, and workload, TUMs preferred more mechanical questions, namely multiple-choice questions rather than skill-based exams. The participants in the first phase seem to be more familiar with summative assessment and multiple-choice exams since most of the time, they previously experienced them. Summative assessment is preferred in Türkiye most probably due to the fact it is an exam-oriented country employing high-stakes testing in the form of traditional multiple-choice exams (Acar-Erdol & Yıldızlı, 2018). Nearly none of TUMs, being inclined to view all of them were the same, could not

differentiate the meanings of the terms testing, assessment, and evaluation. Therefore, as experts voiced in the focus group meeting, at first comprehensive planning about language assessment and testing should be organised to enhance the awareness. In addition, there is a need to train instructors or prospective teachers to ensure their comprehension of the differences among testing, assessment, and evaluation in the first place. Assessment must be recognised to have other functions such as improving learners' performance and giving feedback about teaching/learning. Besides, the perspectives of learners and instructors should be changed to keep up with the recent conceptions of assessment as claimed by Fulcher (2012), Hidri (2018), Scarino (2013), and Tsagari and Vogt (2017).

Overuse of test-based exams seems to have contributed to over-reliance on curriculum by instructors while teaching. It was criticised that language instructors did not teach skills as TUMs did not test skills. That note-taking has been removed in listening exams can be a good example. The justification was that it was above the level of their students so they removed that part even from their curriculum and instructors were told not to teach note-taking in their courses. Most of the TUMs mentioned that assessment is separated from teaching in the practicum, which is against the interconnection between assessment and teaching. Thus, testing is believed to remain at the theoretical base, abstract, and isolated from the teaching counterpart, which might challenge TUMs to put their knowledge into practice. On the other hand, the experts suggested that there is a close relationship between teaching and testing acknowledging that assessment is as important as teaching. This finding supports the results of the study carried out by Bayat and Rezaei (2015). In fact, it is an effective tool not only for learning but also teaching to follow students' progress, the missing points, providing feedback on whether they have learned or not, which shows parallelism with several studies (Hidri, 2018; Rogler, 2014; Yastıbaş & Takkaç, 2018).

One reason for students' less attention to skills might stem from the fact that they are not assessed in proficiency exams in many EPPs. Due to a national-level washback effect, there is a lack of testing of four skills might lead to less focus on teaching and learning these skills. Therefore, the experts in the field put forward that skills must be included within exit examinations as a criterion, which would ensure that they would be taught effectively. This finding is harmonious with Karcı-Aktas and Gündogdu (2020) that maintain that four skills are needed in the evaluation ingredients

in exit exams with established standards. A need emerges to change the testing-oriented mentality into a more assessment-oriented one by updating national-level practices making them more practice-based. The findings from the first phase reveal that insufficient background of the students coming from K12 and test-oriented language education appear to result in that production skills are rarely or in some institutions not included in both tests and language education as a reflection result. It was complained that the production part was problematic not only in state universities but also in foundation universities that had not completed their development. The main problem was believed to be the washback effect of central exams administered by ÖSYM. Using the metaphor “dynamite,” a testing member mentioned that these exams were threats to universities and language education in general. It was explained that when students had difficulties in the exams prepared by TUMs, they immediately took YDS or YÖKDİL since there was no speaking, no writing, and no production-oriented measurement. In return, it seems that production skills are considered a waste of time by students, which causes universities to remove production skills from their exams after a while.

Removal of production skills from classroom teaching appears to lead to a content-heavy curriculum. The TUMs stressed that courses were based on traditional grammar education, which made their exams traditional, as well. While foreign language learning and exams in the world are observed to focus on 21st-century skills, communication skills, and production skills, TUMs at EPPs ask if there is the suffix “s” after third-person pronouns. The exams are curriculum-oriented, just like language teaching and learning process in the classrooms. It is perceived that English is taught and tested through rules even at the university level, but language skills within 21st-century skills cannot be transferred to communication and daily life, which is believed to result from rarely assessing transferable language skills. A TUM, for instance, explained the issue claiming that they generally tested grammar rules they had taught or evaluated if they memorized them. As a result, it was criticised that preparing a test suitable for students’ appropriate language level was almost impossible due to grammar-based testing instead of 21st-century skills. TUMs were admitting that they had concerns over whether their exams were appropriate or not. It was interesting that in some universities TUMs were even not permitted to change topic variety. Due to students’ low level, instructors were unable to bring different subjects into the classrooms, which is considered to result from insufficient K12 language education.

TUMs complained their students could not come up with supporting ideas, major ideas, or minor ideas even though they brought up-to-date topics to the classroom. They were claimed to be willing to end their turn by saying one or two sentences.

Accordingly, the experts proposed that a level-based system should be adopted. It is felt that language education and testing are less test-oriented in level-based system. In addition, it was suggested that philology students should be separated from the general group in terms of the balance in the classroom. That is why the level difference between philology students and others seems to be very high, which might harm in-class level harmony and create different profiles. Also, it was noted that exams should be prepared regarding 21st-century skills. They agreed that transferring communication skills and other 21st-century skills into exams would ensure that students became aware of their communication skills. To make a smooth transition, it was offered that TUMs could form collaboration through social media in which all volunteer TUMs could be added and share their good implementations.

Apart from the issues mentioned above, students learn from previous students, social media applications, or different sources that they can pass the preparation year more easily with the exams held by ÖSYM. Then, they start to make some absences at first and do not attend classes from the second semester. Some TUMs explained that they wanted to increase the equivalence cutting point of these exams, but they could not get permission from their Rectorate. Thus, the syllabus is changed into more testing-oriented. In addition, other TUMs argued that more emphasis was given to testing grammar, vocabulary, and receptive skills than productive skills. This is partly due to the fact that they had difficulties in administering speaking and writing tests. This is consistent with a proposition made by Mede and Atay (2017), that is, listening and speaking were not tested in EFL contexts in an appropriate manner. Likewise, most novice TUMs were found to employ language tests to measure grammatical and lexical points rather than testing skills such as writing and speaking as also approved in several studies (Martinez-Duboc, 2009; Semiz & Odabaş, 2016; Tsagari & Vogt, 2017).

It was criticised in the interviews that there was a heavy emphasis on grammar-based instruction. Education with too much grammar focus seems to lead to difficulty in teaching and learning skills, especially speaking. Due to this focus, students were claimed to be having the fear of making mistakes, which was viewed as a major barrier to their progress. The experts stated that there should be more focus on productive skills

instead of grammar. This finding is also supported by (Yaman, 2018) who states that due to a traditional grammar and vocabulary-based approach at EPPs, students lack communicative abilities. In Türkiye, it is obvious that a person who knows a foreign language, especially English as it is *lingua franca*, is a distinct advantage in finding a better job or having better standards (Arik, 2020). However, it is indicated that the required emphasis was not put on productive skills, local and contextual needs, interpretation, and evaluation issues. Discreet testing is preferred instead of the integrated testing of skills. The experts proposed that there was a need to revise language testing practices in order to provide better training for prospective teachers and learners as discussed in the literature (Berry & O’Sullivan, 2016; Christoforidou & Kyriakides, 2021; Giraldo, 2018; Gotch & French, 2014).

In addition, most of the participants argued that students starting their preparation year had very low English levels, which was believed to influence not only classroom practices and testing. During the interviews, it was frequently claimed that most preparatory students were at the A1 level although they were supposed to have received English for many hours during their primary, secondary, and high school. For instance, one testing member shared the result of their exemption exam at the beginning of the year as 700-800 of the 1.200 students performed A1. These low performances are in line with the results of the British Council report and the EFI results. What is worse, students were criticised to have limited critical thinking or cognitive-cultural knowledge. It seems that since students coming from K12 are rote learners, their ability to produce strategies is very poor. One director explained when they failed, students claimed that exams were difficult although they could not even realise the plot of a text in a reading exam. Due to the weakness of the students' critical thinking skills, TUMs complained that they could not prepare modern exams that are aimed to measure four skills instead of grammar-oriented or test-oriented exams. This is thought to affect their interest and motivation negatively. This can be accounted for insufficient non-class activities. Several TUMs claimed that even in the classroom, their students were insensitive to current news or events. They commented that choosing options in test-based exams seemed to have affected their cognitive levels. In addition to their limited background, they were criticised to have difficulty in the internalisation of productive skills and concepts. This might have resulted from the overuse of tests in EPPs. That is why a number of TUs removed the speaking part from their exams and included more

test questions considering that students had problems with productive skills. In this sense, the absence of productive skills in proficiency exams seems to affect the courses as instructors spend less time on productive skills taking into account the exam content.

Regarding students' low level, the experts proposed that an orientation programme should be given to newcomers about the importance of language learning in their job-seeking period and career. In his study, Z. Öztürk (2015) displays how learning English is useful in finding jobs more easily and career planning. Besides, TUMs were suggested to prepare authentic and up-to-date exams to address the needs, expectations, and interests of the students. Chan (2020) stresses the use of both authentic tasks and authentic texts in testing.

Considering all the issues mentioned above, the findings of the study revealed that washback is a complex, intriguing multi-dimensional issue which should be debated from various starting points.

5.3. Key Finding 3: Standardisation

The third concept derived from the findings of the study dwells on standardisation issues between universities at regional and national levels. As mentioned earlier, there are 207 universities in THE system, most of which provide preparatory education. It would be acceptable to conceive that there are differences and variations in different dimensions or applications ranging from education quality, principles, priorities to testing practices, as well. However, the study revealed that the range of the differences was great from one university to another, even between two universities in the same region. In international level, let us have a look at Europe. The Commission for Cultural Cooperation of the Council of Europe, started the work of the Common European Framework of Reference for Languages (CEFR) in 1971 in order to bring the member countries closer to each other and ensure that the languages of the member countries could be taught to each other more systematically and published its final version in 2001 as the result of many studies and communications. While unity has been targeted in language skills in the European Union, universities in our country appear to diverge from each other having more different education quality, system, implementations, which is alleged to result in inequalities amongst universities. For instance, in language assessment practices, TUMs at one university prepare mainly

grammar-oriented, test-based exams, but others prepare completely skills-based exams. Or it was found that some universities placed less emphasis on productive skills, focusing on grammatical issues while others placed more emphasis on communicative skills of the students, integrating use of language into skills within context. For Eignor (2001), standardisation helps provide criteria for testing practices and the effects of test use. Most directors complained during the interviews that a student exited as B1 from a university, but as A2 from another university, or as B2 from the other one, which clearly shows that different universities have different results. As a solution to this challenge, the experts suggested that the structure of TUs should not be left to the university and school directors, standard practices and minimum requirements should be determined to cover all universities, but taking into account of differentiation due to needs, aims and infrastructure of the institutions. In addition, Linn (2006) proposes that standards could be employed as a checklist by experts. Besides, the experts stressed the importance of benchmarking studies between universities. One director, for example, explained that benchmarking studies on national and international exams would contribute to the evaluation of language assessment practices of universities. In this regard, it can be concluded that it is crucial to set a standard framework for the assessment practices. It can be supposed that through benchmarking attempts and processes, universities would learn good testing practices, assess their operations and results and collaborate and cooperate to better implement the language testing process. In the absence of a national framework, the institutions are found to adopt various scales, grading, cutting points, testing system, etc.

A possible explanation for why the challenges appear to fluctuate and be diversified around Türkiye may relate to the constant system changes as the decisions are taken immediately and leave their place to instability at a sudden. During the interviews with the directors and TUMs, the lack of legal ground was frequently mentioned in that since testing units do not have a legal place within the organisation, TUs can be affected by changes at university, school and unit level. Some elaborated on the lack of a settled system stating that the school director and/or deputy directors changed when the university administrators changed, which seemed to change testing practices and processes ranging from book selection, exam format, exit criteria to question items. Universities have been transformed by going through significant changes in terms of structure, hierarchy, vision, values, and practices (Polster, 2016;

Urbina-Garcia, 2020). Some of the participants complained about the change in books, test system, contents of the exam, calculations, and so on every year. The constant changes were found to bring pressures, overwhelming workload, heavy administrative burden, unsatisfactory leadership, and management. However, the universities in Türkiye seem to go through such changes not only for transformation to keep up with global standards but with each new rector. This may have a harmful long-term impact on the standardisation of universities and may even serve to increase the differences from high-level to low-level quality and results (Franco-Santos & Doherty, 2017; Kinman & Johnson, 2019). In this sense, it would be argued that Turkish universities move away from each other in terms of quality and success in language testing though there is a reverse attempt in European Union to bring countries together through a common language mission and CEFR in detail. In the context of Türkiye, the universities range enormously from the ones employing only multiple-choice questions as YDS to the ones paying attention to skill-based language testing. The emerging question is whether we should support the standardisation of universities in language testing or isolation of them with their unique characteristics.

Due to the lack of standardisation along with other challenges, TUs were found to have ambiguity in the operational system. Some TUMs criticised that they could not establish their workflow to set up a system due to constant changes, which can be a factor making them tired of uncertainty before a flowing pace. The challenges regarding standardisation are likely to result in bigger issues in recognition of language education of EPPs in the same country as an exit level, B1, certified by one university seems to be denied in another university in the same city or region. An important point revealed during the interviews was that lack of standardisation and constant system changes lead to ambiguity of the operations in the language testing process in addition to other factors. This finding is quite logical and understandable within the proposal that structure, operations, and minimum criteria of testing units should not be left to school directors or university leaders. It is noted in the meetings that even the best implementations can be worsened with leader change. This finding is harmonious with Drew (2010) that maintains that there are potential risks with every change decision if change-readiness is internalised within the university.

Another key issue is about quality studies which are believed to be at the surface level with more focus on documentation. Some TUMs whose universities were under

review of YÖKAK complained about the heavy workload of documentation and sudden requests, affecting their planning and workflow. It is challenging to create quality assurance unless the institution and its stakeholders adopt the concept of quality assurance and a quality culture. It can be concluded that the quality culture should be internalised rather than obtaining a quality certificate, otherwise, the work and processes remain as formality. The experts had a consensus on the establishment of a joint recognition between universities. In addition, the EPP director and at least a member from the TU were suggested to be included in university visits by YÖKAK. When there is a lack of established quality process, the attention is drawn to the internal processes of EPP, especially the TU. However, most TUMs mentioned that they had serious problems in internal organisation such as insufficient coordination among the units and ineffective coordination among TUMs. For instance, a TUM admitted that they could not always help or consult each other to exchange ideas while acknowledging that the importance of coordination in exam preparation, implementation and even following processes. Another TUM added that directors were changing, testing members were changing, but there was one thing which did not change and that was the problems they experienced. Considering irregular, unregistered, undocumented exams, it can be understood that previous testing members only save the day while the newly arrived team is like looking for fish in the ocean. A TUM working at a state university complained about the lack of regulations by national authorities, stating that they could not develop a system under current conditions at the unit level. The experts proposed that a consortium or body institution could be established between public and private universities considering the geological and institutional proximity. Additionally, it was suggested that a national framework on language qualifications including common minimum criteria, exit scores should be determined by relevant authority councils.

Local needs or contextual factors such as university type play a crucial role; therefore, TUMs should be competent in carrying out appropriate assessment processes in their universities, which requires them to be assessment literate and to pay attention to the contextual needs in assessment. However, most TUMs in state universities were found to be focusing on knowledge and grammatical structures and lexical points. TUMs in some universities thought that existing implementation was sufficient while others tried to update their processes by adding new topics such as alternative assessment, dynamic assessment, computer-based testing, and statistics to their existing

implementation. In this regard, according to the experts, TUMs might cooperate with other universities in order to share good examples regarding examination processes. Furthermore, it was also noted that a common exam calendar valid for different units or student groups should be prepared.

Most directors argued that one major challenge was related to differences in the exit criteria. They explained that to a great extent, targeted exit levels vary among higher education institutions. In other words, EPPs have inequalities in exit levels as the skills assessed vary considerably among universities. This challenge is considered to result in a greater problem that a university does not acknowledge the successful completion of students at another university. In the report by British Council (2015), universities which prepare comprehensive and valid examinations were advised to share their samples as models with other developing universities. Nonetheless, as mentioned in the methodology of this study, many universities refrain from sharing information about their examinations or sample tests. Even though quality assurance initiatives by DEDAK with the aim of standardisation in EPPs started in 2012, they have not been able to provide solutions to date.

The Council of Higher Education maintained that at least B1 is determined as the exit level for Bachelor's degree students, while the report by the British Council suggested the exit level for students EPPs needs should be revised as B2+ for philology departments and B1+ for others. In addition, experts highly suggest the balance between students' number and language instructors at EPPs. Otherwise, it leads to an extra workload for limited human capacity. For instance, the reason for the elimination of standardisation meetings is possibly due to the workload of both TUMs and other instructors in EPPs.

The issue of reliability emerged as a major challenge even during the scoring of writing quizzes. The TUMs complained about maintaining the standardisation among raters. It is difficult to succeed in rater reliability without in-depth training. The issue is related with the extent to which different raters are in agreement with each other on the scores (Fulcher, 2014). However, introducing raters into the assessment procedure is problematic due to their subjectivity inherent in language assessment (McNamara, 2000; Szmerdt-Chandler, 2011). Therefore, subjectivity in rater judgment(s) is one of the major threats to the reliability of test scores (Bachman & Palmer, 2022; Bachman, Lynch & Mason, 1995; McNamara et al., 2019). One factor hidden behind fluctuating

interpretations of scales may be that band descriptors are vague (Huang et al., 2018). Comments shared during the interviews illustrated the situation fairly well. Their accounts reflect avoiding inconsistent scoring due to the subjectivity of raters is almost impossible without training and fostering instructors. Thus, the quality of rater scoring has a significant impact on test reliability. Yet EPPs cannot make the instructors receive the necessary training due to their financial limitations. Then the issue regarding lack of training seems to have a domino effect on other challenges within test development and evaluation processes.

What is worse, the instructors working at EPPs are not regarded to have enough experience, nor sufficient training in language testing, which causes more serious challenges, including their loss of motivation and reluctance to work in these offices. This lack of training will be explained in detail in the following part.

5.4. Key Finding 4: Training/Specialisation

The fourth key concept is training and specialisation. Lack of pre-service/in-service training was found to be a major deficiency though it is indicated that all the instructors and potential testing members are expected to know how to prepare language tests regarding their undergraduate education or pedagogy training. Yet in real terms, it can be challenging for novice instructors or even experienced ones to work on designing language tests unless they have special experience and expertise (Téllez & Mosqueda, 2015). In addition, it is difficult to find faculty staff specialized in language testing who may provide training at each university. Therefore, most of the participants pointed out they tried to persuade publishers to invite specialists to provide certain training on item writing, item analysis, and so on. It was also noted that the universities which did not have such a chance had changed the direction of their exams as the earlier tests would question functional and practical knowledge more; however, the newer ones consist of more structured and standard-like multiple-choice questions. For Hatipoğlu (2015), it is evident that exams have serious reliability and validity problems and that in-service training is needed for testing and evaluation. Not only in language testing and assessment, but there is also insufficient training in the practicum schools, which was found the main reason for student teachers not to utilize technology (Merç, 2015). This insufficient training may influence the limited use of technology in language testing

processes at EPPs. In addition, based on the data, it is apparent that the rate of pre-service training seems to be quite low due to a number of underlying reasons such as financial matters of the schools, the limited number of experts, the low awareness level of the need, etc. In essence, there exists a struggle against time and deadlines. In my opinion, this shows that testing members seem to complete tasks for each exam in an endless cycle as one exam follows the other without or with limited knowledge. This can be justified by the very rare orientations within the unit, as claimed by the participants that due to time they cannot organize orientation meetings for the starters. Framed from this perspective and taking the comments of the experts in the focus group into account, it might be suggested that especially pre-service training supported by in-service ones at certain intervals be organized for testing members. This result is in congruence with the findings of some research studies in the literature (Mede & Atay, 2017; Ölmezer-Öztürk & Aydın, 2019; Sevimel-Şahin & Subaşı, 2019; Şişman & Büyükkarcı, 2019; Vogt et al., 2020). Experts took this point one step further by also suggesting that geopolitically, at the national level, consortiums of universities or body institutions, one state and one foundation, could be initiated so that good practices can be shared and mutual action can be taken against challenges.

Most of the TUMs discussed some of their colleagues received a limited education in language assessment and testing. Even there were some who did not take any training on how to develop language tests or assess their students. This result is in congruence with the findings of some research studies in the literature (Gebril, 2017; Giraldo-Aristizábal, 2018; Hakim, 2015), in that the ones either having basic training or having no training in language testing were unable to deal with language testing properly; while turning into traditional tests. In addition, it is argued that their teaching and testing did not overlap. Indeed, in-service training is supposed to be beneficial for the TUMs who already received training to refresh their knowledge gaining new assessment skills and learning from other TUMs' experiences. Through well-planned training at certain intervals, TUMs would become more aware of language testing, learn basic concepts, principles, and types if they are novice, and be able to develop, implement, adapt and evaluate language tests in terms of language skills and areas. Similarly, the training would ensure TUMs are more aware, familiar with basic concepts, critical about testing, acquire the ability to select and write appropriate test items, and feel more competent. In this way, they can obtain information, knowledge,

and skills about criteria of testing such as reliability, validity, table of specifications, etc., types and purposes of tests, and how to plan, construct and apply appropriate language tests for each language skill and area.

The findings of the interviews showed that close-ended questions were preferred in the assessment of receptive skills. Therefore, TUMs prepare answer keys for multiple-choice or true-false questions in general. However, there are different types of rating scales employed in the assessment of productive skills (e.g., holistic, analytic). A rating scale is intended to guide the rating process (Wind & Peterson, 2018). The selection of the rating scale might also affect the results of inter-rater reliability. The analytic scale is generally preferred due to the fact that “more scores or pieces of information about a test taker’s ability will translate to greater consistency” (N. T. Carr, 2011, p. 133). In addition, scoring different tasks separately might result in better reliability since it ensures greater control over rater behaviours, minimizing the subjectivity of raters.

As explained in detail in the previous part, reliability and standardisation among raters are rather challenging. However, even when scales are inadequately determined, raters who have been trained in the use of scales might deal with the situation without any trouble. During training sessions, raters are able to internalise a common understanding of the scale descriptors and apply them consistently in operational speaking tests (Fulcher, 2014). In this sense, specialisation through training can be perceived as the means to achieve greater reliability of scores produced in language tests.

Some have taken undergraduate courses or attended a few seminar courses. However, further training is necessary to help them gain practical skills while being engaged in English language testing and assessment. The progress of TUMs needs to be ensured. However, they admit that they lack knowledge of other features of testing and assessment including the rationale behind choices, ethics, and fairness, statistics of the results, contextual or instructional needs, and social concepts related to assessment (Fulcher, 2012; Giraldo, 2018; Stabler-Havener, 2018). Therefore, they must acquire the necessary knowledge and skills of language assessment, being well aware of how to assess as a part of their professional competence (Büyükkarcı, 2016; Sevimel-Şahin & Subaşı, 2019; Şişman & Büyükkarcı, 2019). Publishing houses organize some seminars and workshops. However, it is stated that they do not see much benefit from one- or

two-day training. Almost all TUMs stated that the content of such events provided generally basic training like an introduction to language testing in terms of its theoretical side, yet it was deprived of practical issues, not improving their testing skills. The TUMs argued that the training provided by publishing houses did not respond to their needs and expectations. They expected more practice than theory in the training was given to them to enhance their practical testing skills since one of the shortcomings was to transform their knowledge into practice. No matter how many years they work, unless they have special training or education, most appear to feel that they were not able to design and administer sound assessments. This finding is consistent with the findings of several studies (Hatipoğlu, 2015; Jin, 2010; Ölmezer-Öztürk & Aydın, 2019). Thus, training with practical applications is needed as early as possible. However, training that is repeated at regular intervals with up-to-date information in line with the expectations of the school, students, and academic staff would be beneficial in changing the attitudes of TUMs and other instructors. This finding is similar to Hildén and Fröjdendahl (2018)'s and Papadima-Sophocleous (2022)'s studies in which a well-organised training programme would ensure a positive change before and after the training.

A number of TUMs indicated that they had great difficulty in evaluation i.e., providing feedback, scoring, and marking due to limited training on analysis and interpretation of test results. TUMs studying at the faculty of education think that the course they have taken only in the last year is insufficient for their undergraduate education. Graduates of other philology departments, on the other hand, criticize that the subject of testing is very superficial in the formation courses. This finding shows parallelism with several studies (Büyükhıskı, 2020; Hatipoğlu, 2010; Sheehan & Munro, 2017; Vogt et al., 2020) which reveals that prospective teachers must have required assessment literacy. Therefore, a separate, intensive, comprehensive, from basic level to detailed, and well-organized training on language testing and assessment is suggested to be provided by experts who are professionally competent in the field of language assessment. In this way, trainees could experience more samples, examples, activities, or tasks especially realistic ones, to improve their practical skills. Similarly, trainers may put more emphasis on post-implementation tasks such as evaluation of performance and interpretation of results. However, training in teaching and assessment should be delivered in an integrated way. Yet, in real terms, just a single session by a

native speaker who works for the relevant publishing house is given in a year or never, which results in a lack of practical issues in a limited time to emphasize each significant topic. What's more, even some terms are not mentioned in the training, at all.

Training is advised to include recent assessment methods such as alternative assessment (e.g., portfolio, self/peer assessment, diary) and computer-based testing, statistical processes, the analysis and interpretation of results, and testing overall language ability. It is argued that pre-service education serves as a significant construct in language teaching. However, when they do not receive necessary pre-service or in-service training, they remain desperate when they encounter unexpected situations, with the possibility of making serious mistakes. This dead-end destiny is thought to affect TUMs' emotional feelings such as loss of motivation whereas an increase in anxiety and giving up due to burn-out syndrome. As crucial, the emotional factors will be touched upon in the following part.

5.5. Key Finding 5: Emotional Factors

The fifth key concept centres around emotional factors including motivation, burn-out, and anxiety of TUMs. Having a look at the issue in detail, it seems that there are four main problems underlying emotional factors: legal loopholes, the structure of testing units, agent-induced, and procedure-based ones. These reasons appear to decrease the motivation of TUMs working in testing units, increase their anxiety levels, and as a result make them feel burn-out. The findings showed that the intense pace of the examination process, the lack of extrinsic motivation, commonly subjective and negative criticism instead of solid ones, and an alert state from the preparation phase to the last part seriously affect their motivation. The motivation of test developers is argued to depend on the importance and the difficulty of the exam (Stansfield & Winke, 2008). Besides, TUMs seem to face a number of issues such as the high number of students, clashes with students, instructors, and directors upon the slightest mistake, and keeping up with time, all of which lead to a deep level of anxiety among TUMs. As a result, the findings of this study support the recommendations and findings of Kartal and Basol (2019) and Oktay (2015). It is also seen that burn-out feeling is common due to the high working tempo with overtime duties for a long time and difficulty in taking

leave. All in all, TUMs working in these units do feel the necessity for rotation in a short time.

Regarding the legal loophole, it seems that a systematic balance cannot be achieved in the workloads because their job descriptions are not legally determined. Apart from that, it is claimed that they sometimes face harsh reactions, even accusations, by students or instructors, yet they cannot defend themselves and remain helpless because of a lack of legal rights. Due to the lack of their job descriptions in legislation, TUMs cannot receive financial compensation for their hard work. However, a certain fee is paid to instructors for each lesson taken after the compulsory hour, but TUMs working in testing units are not paid any fee even though they take uncompleted work home following office hours. Likewise, the compulsory course hours of the director and vice directors are reduced due to their duties, yet TUMs may not get such an opportunity under current legal conditions; thus, TUMs are required to fulfil their compulsory course hours while having to work at the office afterward, which decrease their intrinsic motivation as defined in the study of (Syamananda, 2017). Due to the legal problem, the instructors do not want to work in these units. An interesting finding is that the testing unit is the least demanded one when instructors have to choose to work in one of the units in the school. In this case, the director forces especially new instructors to work in the testing unit and assigns them an official letter as *fait accompli*. There are various reasons, such as excessive workload, the difficulty of the job, time problems, blames, and clash, sacrificing private life, why this unit is not attractive. In fact, directors generally make an announcement for the missing position in the unit, but when there is no application, an instructor as a victim is identified and forced to work either by persuasion or by forced assignment.

At this point, there is an important difference between state and foundation universities. Penal sanctions are made very indirectly when the selected person cannot prepare an exam with sufficient quality in state universities, but instructors may be threatened with job security in some foundation universities. TUMs, who do not want to risk their job, strives to show their best and improve themselves with their own efforts. TUMs who are found to be underperforming at public universities can be taken out of the office as a direct penalty, yet in this case, it is more difficult to find someone else, so the principal tries to persuade them to continue to work. Considering the above-mentioned reasons, testing member turnover is excessive in that even newcomers start

to count the remaining days of freedom to leave the unit. It seems that TUMs do not consider themselves a permanent expert in the unit, which may also block specialisation or training as directors do not want to send their staff to a long-term training programme since that member would desire to leave the unit in a short time. As a result, testing units seem to be understaffed as the number of members decreases year by year. One member alone might prepare a test for over 1500 students being responsible for the whole level, which appears to put test security at risk in the case of illness or unexpected situations. For experts, under the coordination of YÖK, a study should be carried out with the participation of preparatory school administrators, and the title of TUM and job descriptions should be determined for those who will work in testing and other units, which should be added to the legal legislation study. It is recommended that the course load of TUMs who work in the units be reduced based on the total workload in the unit while the legal work is being done. Considering that there is a cycle as one task is completed, the next one starts, the average course load of TUMs should not exceed five-six hours as applied to the administrators. An interesting finding is that all the experts agreed it would be inefficient not to give course to TUMs, but the limit should be determined. It is thought that attending the course at a certain rate provides benefits such as not being separated from students, looking more consciously and holistically, and being more successful in pacing the level.

An important difference between instructors and TUMs is that instructors can go home at around 13:00 after the class, but those working in the unit are expected to complete office hours and work until 5-6 pm. It is stated that when the office is closed due to force majeure, complaints are made by the instructors and students. The more interesting part is about summer vacation planning. In most state universities, instructors do not have to come to the university without the need to take annual leave if they are not going to leave the city in the summer, but those who work in the unit are expected to come to the office like administrative staff and plan their vacation in advance as they will prepare for the new year. Although the pandemic conditions bring the opportunity to work from home, they are requested to return to the office again with the decrease of the effect of the pandemic. Despite the expectation to work in the office, it is unfortunate that the physical conditions in the office are insufficient. Especially in state universities, testing units are given small, airless, dusty rooms on the ground floor or basement floor. The presence of devices such as photocopiers, optical readers, as

well as resources and old exam documents in the room causes a suffocating atmosphere. Considering that someone who has to work in this environment has asthma, he would be constantly using drops, of antihistamines. Although they have to work in the office, due to insufficient technical equipment, two or three people can be given a computer and employees are expected to carry their own laptops.

In addition to insufficient physical conditions, there is ambiguity in the operational system. It was stated that an uncertain tempo and rush situation occurs in the unit when the workflow is not clear. In this case, the employees make an effort to save the day. Whereas, it seems that TUMs need a comfortable work environment with the necessary space and equipment. The complexity of the room where TUMs are located causes many additional problems such as decreased motivation and loss of concentration. As there are not enough lockers, exam papers and packs are piled on the floor. It is also seen that there is endless traffic in the office. While one TUM is separating the envelopes, another is making photocopies, and a third is preparing or trying to prepare exam questions under noisy conditions. The traffic coming from outside does not stop as one instructor comes to leave a paper, another asks for the class list, students come and make their complaints from time to time. How can they concentrate on the exam preparation when the workflow is not clear in this traffic? According to the experts, a large-scale needs analysis should be made regarding physical, technical, and human deficiencies and the results of the analysis should be reported to the authority council, namely YÖK. After a detailed examination covering different universities, the minimum conditions for the continuity of the preparatory schools can be determined and distributed to the universities in an official letter. Official letters from a higher-level body in hierarchical terms can be followed more carefully by university leaders. In addition, it is recommended to bring flexibility in home office work of TUMs, albeit limited, in universities where physical conditions are insufficient.

Most of the participants voiced that the topics were very theoretical and abstract for the students. However, there was no time to update topics, improve their testing abilities, or practice new issues in their development processes. In addition, it can be inferred that the knowledge of directors in state universities was at the basic level because they were appointed from different fields rather than at least from the philology department. These managers were focusing on the realisation of an exam, disregarding

its key function. They were even unaware of basic concepts, principles, techniques, and approaches of language testing, types, purposes according to different language skills and areas, testing criteria, reliability and validity, and interpreting of the results.

Another major challenge results from attitudes and behaviours including students, instructors, and directors. Students who come from K12 at an insufficient level may not be able to reach the desired target level in a short period of eight months. After a while, it is observed that attitudes of disinterest emerge in these students. Their indifference to the lessons is also reflected in the exams, as they come late, forget their identity cards, sleep in the exam, blank papers, we even force their names to be written.

One of the reasons behind these attitudes was the dominant misbelief that students can pass the preparatory class without much effort or the influence of ÖSYM exams such as YDS and YÖKDİL. On the other hand, it is stated that there are also students who complain and cause uneasiness when the grades are not announced as soon as the exam is over. When we look at the reasons, it is seen that the students want to make a complaint with subjective and unjustified arguments such as the exam was difficult in general, the teacher of the other class gave higher.

According to the findings of the interviews, it is emphasized that most of the complaints come from students who attend classes irregularly and have low performance in the classroom. However, it is important that their scores can be lowered in the re-evaluation of their results due to their objection. Nonetheless, it is stated that students tend to object genetically and intrinsically, and a result-oriented generation is raised as a society. An important problem with complaints is finding the exam documents of the student from the previous packs and re-evaluating them and writing a report. Problems such as space problems in testing units, lack of cabinets, and the lack of a proper archive can cause loss of time, energy, and motivation.

In addition to students, instructors' low work discipline and irresponsibility have a negative impact on TUMs. They can fall behind the programme in the classroom. Some instructors also report 1-2 days before the exam that something unpredictable has happened. Our instructors can delay their tasks such as reading, handing over papers, and entering scores, bring the papers in an unordered way, and forget necessary documents; i.e., signature list. They don't even follow the directives, instructions, and checklists.

One of the reasons behind the irresponsible behavior of some teachers may be that our professors usually do extra work with a private institution or publishing house and transfer their concentration there as they earn additional income. It was found that directors from other disciplines than ELT or philology departments may be inattentive to the issues or challenges due to a lack of relevant knowledge. Neither do they want to deal with problems, nor do they want to make a change, which means that no action is taken. What is worse, considering the fact that the failure and negative attention are focused on TUMs. Complaints about exams can be made by students or teachers to the administration. In this case, without even making an explanation, the manager can be influenced and give undesirable bad reactions. In other words, TUMs are standing on a target board.

In addition, distraction caused by instructors seems to result from their lack of empathy. TUMs, who are trying to maintain their concentration to prepare the exam, may be distracted by the instructors who suddenly come to the room, ask about the rules although written in the instructions, and claim that they need detailed explanations in the middle of the concentration, which may risk test security as there may be exam paper drafts on desks. They note that some instructors do not consider their situation, how hard it may be to prepare a sound test. Bearing in mind that exam preparation is not an easy task and requires intense attention, TUMs may be distracted too much during the exam processes with unnecessary explanations to students and instructors, and thus should work in a more isolated way. The experts in the second phase suggest that the director be appointed from ELT or relevant philology departments, which would increase the possibility of noticing the difficulties experienced and contribute to the improvement of processes.

Furthermore, it is also advised that working in testing units may lead to higher-level appointments such as director or vice director, which is likely to prioritise the service in these units. Another crucial solution suggestion is about the management of the relations of TUMs with instructors and students. It is recommended that the relations of TUMs with instructors and students should be maintained at the management level, that is under the leadership of the director/vice director. Otherwise, once they comment on the exams disrespectfully, TUMs may feel that no respect is shown for their work and effort. Besides, assuming that TUMs may sometimes be emotional, the issue will get worse like constant clashes and conflict when they show

their reactions directly to TUMs, which is likely to decrease their motivation. In order to ensure their concentration and motivation, a free room might be arranged for meetings/group work so that other TUMs do not get affected and test security is maintained when a student or instructor comes to the office for inquiry.

5.6. Key Finding 6: Technical/Physical/Human Infrastructure

The sixth finding is focused on infrastructure necessity on technical, physical, and human capacity in testing units. The main problem that emerged in the first phase of the study dwells on the problems related to technical equipment. Major technical problems are the need for new equipment and the constant failure of old devices. TUMs claimed that there might be problems with listening, the smartboard, and low sound quality. It was also stated that projectors, speakers, copy machine, and optical reader may be old and not working if exist. Considering the fact that technology is something that wears out quickly, the old equipment needs to be constantly renewed or repaired. Since the budget of EPPs is very limited, they cannot replace the old ones easily and quickly; therefore, the service is being called, but it takes a long time for the service to arrive and they lose time. When the needed devices cannot be provided, they have to buy them by their own means to solve the problem or they may experience problems due to the lack of these devices. For example, it has been widely stated that licensed programmes are needed in the preparation of the exam, in making changes to the text, in statistical analysis, in listening and in making arrangements, but that they cannot be provided by the university and the school. In this case, TUMs can pay the license fee themselves, are forced to use pirated software, or not use them at all and prepare the exams as in the old system. Frankly, either case does not seem like a suitable method. In a more interesting scenario, when TUMs save money and install the individual membership on a single machine, the office turns into an internet cafe and the teachers queue to use that programme.

On the other hand, in terms of test security and risk management, it was found that some devices such as surge protectors, hard disks, cameras, and steel cases are needed. For example, TUMs seem to be desperate about what to do if the computer breaks down as they do not have a hard disk to back up and store exams against system failure. It was sad to hear from the participants that sometimes the only thing they can

do is pray for God. Besides, it was commonly explained that the automation system for grading developed by a company through outsourcing in most universities seems to be incompatible and fall short in calculating the scores of the preparatory programmes such as quizzes, midterms, out-of-class or in-class activities. Thus, TUMs are required to download all the scores from the system and enter score calculations manually, which leads to a waste of their limited time. According to the experts, the EPPs, and testing units in specific terms, should be supported by the university in a systematic way, not based on the temporary leaders. It was highly recommended that the necessary technical equipment and software programmes be supplied based on the needs. It was also suggested that under the coordination of the authority council, a consortium of IT departments of some universities could work on a more appropriate automation system that includes the differences between EPPs.

According to Fulcher (2014), the use of double rating as blind scoring would avoid the potential influence of a rater on test scores. However, in most of the universities interviewed, the state ones in particular, only one reading is conducted due to the insufficient number of instructors and their heavy workload.

5.7. Key Finding 7: Sustainability/Efficiency

The seventh finding is related to the sustainability and efficiency of testing units and the operations to be conducted by TUMs. Due to the lack of legal ground of these units, there is a constant change in the operations and staff. As mentioned earlier, there is no such unit and the title testing member on the legal ground. Thus, the structure and operations in these units seem to be changed once the director of the school changes. Because of excessive workload and lack of extrinsic motivation and other factors, TUMs do not want to work for a long time. Hence, instability cannot be maintained in testing units. Another result drawn from the study is that due to the challenges, most state universities seem to change their exam format in midterm, final, and even proficiency to multiple-choice tests, which is highly criticised to be employed in YDS and YÖKDİL. The experts recommend that the legal ground of testing units, the title testing member and their workload be established as a result of a detailed study and needs analysis under the coordination of the authority council. It is also suggested that a national umbrella framework, namely Higher Education Language Qualifications

Framework, should be established by the authority council, which ensures that exit levels need to be revised. Besides, key activities should be arranged to enhance the awareness of language assessment among directors.

Administrator change appears to have a negative impact on the system of testing units. Especially the new director who is from other disciplines than ELT or philology departments seems to start to make a change from these units questioning the exams. As a result of their lack of testing knowledge, the failure and attention are focused on TUMs determined as a target. Emphasizing the importance of field expertise, some TUMs note that better decisions can be made if the right people from the relevant field are appointed to their school. Otherwise, the directors from other disciplines do not or cannot deal with their problems either lack of knowledge or interest. In this unsystematic structure, there is insufficient coordination between the units with no or limited consultation with each other. In addition, within the unit, there is ineffective coordination among testing members although it is very important in the preparation, implementation and even following processes of the exam as TUMs in some universities prefer to work rather individually. It is believed that once legal studies are conducted, the minimum standards and quality could be set. Another key solution could be to develop a physical and online archive system.

Since there is excessive testing staff turnover, sustainability and specialisation are not achieved in these units. Due to the fact that exam preparation itself is a stressful, difficult process and there is no additional payment, neither is everyone willing to work nor the ones selected and working want to stay in the unit. The challenges experienced during the previous team seem to continue in the following group of TUMs as a dead-end cycle. Let us take two key processes: archiving system and prior documentation. In terms of archiving, there are two major challenges: space problems and extra workload. Since they do not have enough space and extra staff, archive work is left aside at these units. Yet once requested, TUMs who archive papers randomly may not be able to find the relevant exam document. Regarding documentation, TUMs complained that document work in the office is too much. Due to the lack of prior documentation, it is challenging to deal with past documents and files while there is a flowing process. It is recommended by the experts in the second phase that task and responsibility sharing should be determined before the term. Lastly, it is suggested that a clear workflow must be prepared to show the timeline of the tasks and responsibilities in each task.

Most TUMs complained that with the aim of providing a more general English at EPPs so that students would be able to communicate in social contexts, they are not able to prepare students for faculties due to a lack of relevance between faculties and EPPs, which is also found in the findings of the report by British Council (2015) in which it was noted that what students learn at EPPs might be irrelevant to the courses in their faculties. In this way, it is thought that there is a lack of connection between faculties and EPPs; however, building bridges between the two both ensures learners' language learning progress and motivates them by showing some examples of content they will study at the relevant faculty. In line with this shortcoming, experts focused on the need for collaboration with faculties in order to identify the characteristics of the department and of the language that is used as the medium of instruction at faculties so that EPPs could meet the needs of faculties. In this vein, common goals and exchange of information can be established for the benefit of students. This finding was also found in Keser and Durmuşoğlu-Köse (2019)'s study in which the students at EPPs may internalise the value of learning English when they can use it for research in their courses at the faculty by reading resources in English during their studies. Accordingly, all experts suggested that the curriculum at EPPs be transformed from General Purpose English into English Academic Purposes or English for Specific Purposes focusing on students' needs and expectations. Yet one challenge in this scenario is that students come from different departments and it would be almost impossible to set up a class for each department. However, it could be recommended that the students could be gathered under the major field in which the students coming from Business Administration may receive language education with the ones from Economics since the two departments are in the same field according to the International Standard Classification of Education by the United Nations Educational, Scientific and Cultural Organisation (<https://uis.unesco.org/>). In this way, field-related content could be integrated to increase students' motivation within EPPs.

5.8. Key Finding 8: Digitalisation

The sixth finding is focused on the digitalisation of some processes and some parts or full parts of tests. Even European Commission prepares to transfer their paperwork to online documentation through the project of Erasmus Without Paper. In each office at universities, a great deal of paperwork is conducted and a lot of papers are thrown away.

One key challenge in digitalisation is the difficulty of adaptation to change. Digital testing can be applied as an automated device for monitoring and checking knowledge by a teacher or a self-monitoring tool that provides oral and written visual control of the results (Baryshnikova, Kostenko & Voskoboynikov, 2021).

Digital transformation is a dynamic process. Testing today is undergoing a transition where testing agencies are designing new models of administration and scoring (Dadey, Lyons & DePascale, 2018; LaFlair et al., 2022; von Davier, 2017). However, it is thought that poor resource planning and lack of IT capacity are some of the key challenges against digital transformation. On the other hand, digital technologies are radically transforming students' habits. Test developers are striving to develop more test-taker-centric designs while maintaining high standards of score reliability and validity of score interpretations.

In this vein, multimedia Internet resources might present interactive and entertaining information through text, animation, graphics, sound, and video. For example, survey monkey is an Internet resource used by both students and teachers to create all kinds of tests, questionnaires, and quizzes. It is comfortable both for performing various tasks using the technology of task-based learning and content-language integrated learning and for receiving feedback from students or a teacher. A resource called Classmarker is similar to it. Thus, it is advised to build more flexible and efficient units where the existing methods are drastically transformed, by providing an appropriate response to the students' needs. TUMs noted that although students and instructors are resistant to change, they are condemned to perish in a figurative sense if they do not change by adapting to new circumstances.

Test security, including the processes and procedures that are developed to prevent fraudulent test-taking practice and its maintenance are required to ensure score integrity. The two most common fraudulent means are believed to be cheating and the

theft of test content (Foster, 2013). Cheating can take on different behaviours, but almost all cheaters aim at obtaining a higher score than the one that would have been achieved through academically honest behaviours. Like cheating, the theft of test content can occur in various ways. The aims associated with theft can be to assist the perpetrator in getting a higher score, to assist a friend, or to seek for a financial benefit by selling test to test takers. In online testing, the most common cheating behaviour is to have a substitute person take the test. For this reason, test taker authentication is critically significant. Authentication contributes to ensuring that the test taker beginning and completing the test is the person whose name matches the identification documents (Foster & Layman, 2013). The prospective test taker has to submit an approved ID with a photo during onboarding. While language testing experts were trying to increase the use of online testing and assessment, it was rather challenging to change students' preferences very easily until COVID-19.

The emergence of COVID-19 announced by the World Health Organisation (WHO, 2020) as pandemic, following its rapid spread across the world, led to an unexpected lockdown in nearly all countries. Like other sectors, higher education institutions were forced to make transition from face-to-face education on the campus to online courses, creating ambiguity and putting a direct impact on all stakeholders in higher education. As an immediate response to COVID-19, higher education institutions were closed in almost all countries and courses were shifted to online to deal with the urgent need of ensuring the progress of education. According to a report published by UNESCO (2020), since the spring of 2020, when the pandemic broke out, face-to-face teaching has been suspended in more than 144 countries, affecting approximately 1.2 billion students all over the world. Furthermore, it was stressed that more than 90% of students across the world were influenced directly or indirectly by school closures. New data reported by UNESCO (2021) illustrates that about 1.5 billion students across different continents have been influenced by the sudden transition, and over 800 million students have faced serious challenges in terms of disruption to their education. According to the report, 31 countries took measure of full closures while 48 countries kept their education through part-time schedules (Erarslan, 2021). This crisis has significantly accelerated the shift toward the online environment (Burgess & Sievertsen, 2020). Meanwhile, it was compulsory for universities to shift to online education by upgrading their technological means. Due to suspension of face-to-face education, e-

learning was left as the only option for teachers to communicate with their students (Abduh, 2021). Therefore, most of those students rely on online courses to continue learning, and many teachers have to cope with the abrupt shift from offline to online teaching (MacIntyre, Gregersen & Mercer, 2020; Moser, Wei & Brenner, 2021).

To sum up, the emergence of the pandemic has made online education and formative assessment more pressing, since many higher education institutions need to switch from face-to-face to online education while online teaching, learning, and assessment have become “the new norm” (Todd, 2020, p. 5). To foster learning-teaching process during the pandemic and beyond, it has been crucial for teachers to engage with online formative assessment practices (Rahim, 2020), and for researchers to examine teachers’ assessment implementations in online contexts (Mimirinis, 2019).

Within the current chapter, the findings of the study were corroborated by the related literature. The discussion of the findings in line with the related literature enabled us to be aware of some long-lasting challenges preventing the effective implementation of language testing. As well as such long-lasting challenges, the issues related to the roles of the stakeholders, policy-level actions, and practical aspects were also pointed out. Based on the discussions of the findings, implications are presented in the following section.

To sum up, most problems emerge as systemic. Although the problems encountered in unit-level language assessment are expected to be highly technical, the findings in my study demonstrate that most of the problems experienced by stakeholders are actually related to human factors. Some of the problems can be solved through training. These problems cannot be solved by simply hiring specialised testing staff members. Rather the systemic inherent in the current practices need to be addressed at a macro or national organisational level.

This section summarized the major findings of the present study while discussing the relevant issues with respect to each research question. The following section will present the conclusions, implications for different stakeholders, and suggestions for further research.

CHAPTER 6

6. CONCLUSION

This chapter begins with a summary of the findings of the present study, followed by the conclusions drawn from the study. First, it presents information about the aim, research questions, methodology, and findings. Then, it presents implications for authority bodies, university/school leaders, and TUMs. Finally, the chapter ends with some suggestions for further studies.

The review of the literature has shown that no study has been conducted about both the challenges and solutions in language testing processes and procedures at EPPs in THE system in a comprehensive way. Therefore, regarding the diversity of universities, EPPs and assessment systems in place, this study is expected to be a significant step towards holistically identifying the common problems and addressing them in a standardized, well-informed, and professional manner.

6.1. Summary of the Study

The present study first aimed to find out the challenges encountered in language testing practices of EPPs in Turkish higher education institutions. Second, it also attempted to search for feasible solutions to the challenges. In line with these aims, the following research questions were formed:

- a) What are the challenges encountered in testing practices?
- b) What are the solutions to the challenges determined?

In order to reach these objectives, a qualitative method design was adopted, and the study was designed to include two phases, namely one for the challenges and the other for solutions. In the first phase, interviews were conducted with directors/vice directors and testing unit members separately in order to find out the challenges in testing practices through a semi-structured interview form. In this study, the questions were written by the researcher depending on the results of the literature review, document analysis and informal meetings, and focus group discussion. In developing the interview form, the researcher followed five steps to enhance its quality. Following the determination of the interest construct, a literature review has been done to check whether a relevant interview form has been previously developed. Following the

literature review, an initial draft of the form was prepared according to the categories of ideas and issues encountered in testing practices. When the item pool was prepared, qualified experts conducted reviews of each item in order to ensure that the items were accurate, grammatically correct and did not have item construction problems. Before administering the interview form on the target population, the items were tested to a small group in the piloting step. At the end of all the steps, the interview form was finalised through necessary revisions made in the light of the suggestions of the experts and participants in the piloting.

In the first phase, the participants (N= 208) were selected from diverse universities determined according to certain parameters such as universities' development levels, legal status, establishment year, holding a special mission, and so on. This diversity increased the strength of the sample and the usefulness of the findings. Almost all of the interviews were conducted face-to-face during site visits to the related university before the breakout of COVID-19. The interviews with directors and/or vice directors lasted almost from 45 to 60 minutes, and the ones with testing unit members lasted around from 60 to 120 minutes. The researcher adopted a data-driven approach in order to allow new codes to emerge from the data and categories or themes to emerge from the codes so that a new theory could be generated based on the perceptions and experiences of the participants interviewed. Accordingly, an inductive approach was employed from particular data to general codes and themes. The Constant Comparison Method was employed in the whole process of data analysis. In vivo and descriptive coding techniques were primarily employed to determine the codes. In the coding process, the researcher used six steps as explained in detail in Chapter 3. The audio recordings of almost all the interviews, taken with the participants' permission, were transcribed through non-verbatim transcription. At the end of the transcription process, a 1.365-minute interview with directors resulted in 663 pages (263.055 words) while a 2.775-minute interview with testing unit members ended up with 975 pages (382.707 words) of single-spaced transcription. Nvivo was employed to allow the researcher to read and review text and categorize it easily for the qualitative report. Member checking was employed for inter-rater reliability to make a great contribution to the accuracy and quality of the participants' comments and reflections. Inter-rater reliability was found as 92%.

In the second phase, solutions were sought with seven field experts through the Delphi method. The participants in this phase were the ones having experience in testing units at EPPs in higher education institutions. Since consensus with sufficient information was reached, three rounds were conducted in this phase. In the first round, the experts were sent the findings of the first phase to allow them to be aware of the challenges encountered in the previous phase. The essence of this round was to elicit initial opinions about common solutions in EPPs through free responses to a broad set of challenges. In the second round, oral Delphi was applied online by means of two techniques: brainstorming and focus group interview. The experts were invited to a zoom meeting. In this round, the ideas were discussed and ranked in order to reach a group consensus. The lists of suggestions were revised, edited, and finalised for overlapping and repetitive suggestions, clarification of vague expression(s), the addition of new items, and deletion of unnecessary one(s) thought to be essential. In the last round, the list of suggestions generated in the previous rounds was turned into questionnaire items so that common ones could be selected. In the open-ended phases, the written data and interview data were analysed with content analysis and coding to determine themes, categories, and emerging codes. With regard to ethical considerations, permissions were obtained at different levels taking institutional review board approval, approval of language school, and individual approvals, namely from the directors and testing unit members before the researcher started to gather data. Beforehand, the researcher received Ethics Committee Approval from Anadolu University.

The following conclusions can be drawn from the study based on the research questions and methodological design of the study. It should be mentioned that the following conclusions are presented in a form of a summary.

It is seen that the problems that arise in the interviews with the managers are mostly on an administrative basis and are very general. This situation shows that in some universities, SFL directors are appointed from outside the ELT or philology departments. In addition, it is an important finding that the managers appointed from the ELT or philology departments have very limited experience and knowledge in testing. In this case, it turns out that although these managers have general knowledge about testing, they do not want to deal with detailed problems. The part that the managers especially emphasized in the interviews is that there are very few people who want to

work in these units, especially in the testing unit. In this case, they emphasized that they resorted to a forced assignment. In this respect, it has been clearly stated that a legal study should be carried out; otherwise, they do not want to take risks. On the other hand, it was stated that the principals who have a close relationship with the rector are a little more fortunate, otherwise they could not get the support of the management, and they even had no say in the decisions taken. An important point is that in this case, they emphasize that they cannot provide in-service or pre-service training to the employees in these units because their financial resources are not sufficient.

The most prominent problem in the interviews with TUMs was the lack of legal basis for this unit and the lack of official recognition of its job descriptions and workloads. For this reason, it was stated that they could not oppose forced assignment, especially in state universities, but there were those who thought of being dismissed by not taking quality exams. They stated that the workload, which is frequently mentioned, is high due to the number of exams and the number of students, because they deal with unnecessary secretarial work during working hours, there is not enough time and they have to carry most of the work home. They emphasize that despite all their efforts, they do not receive financial or other rewards, on the contrary, they are on the target board of both students and instructors.

Although the problems encountered in unit-level language assessment are expected to be highly technical, the findings in my study demonstrate that most of the problems experienced by stakeholders are actually related to human factors. On the contrary, those problems can be solved through training, most problems emerge as systemic. These problems cannot be solved by simply hiring specialised testing staff members, rather the systemic inherent in the current practices need to be addressed at a macro or national organisational level.

To sum up, the findings revealed that one challenge seemed to be a cause or effect of the other within the process. Therefore, although TUMs try hard to survive under hard conditions, they cannot fix the problems, but lose time, energy and motivation. During the interviews, some directors appeared to help TUMs, however, they also complained that the issues are more complex than they seem. In line with the findings and discussion of this study, Table 6.1 presents connection between major challenges and solutions regarding the key concepts drawn within the study.

Table 6.1. *Summary of the challenges and solutions with regard to key concepts*

Key Concept	Challenges	Solutions
Legal ground	Lack of job description, title	Establishing the legal ground
	Limited decision making	Defining TUM title and job description
	Limited financial contribution/support	Financial and non-financial incentives by YÖK and TÜBİTAK
	Uncompensated extra work	Authorising TU for complaints/objections
	Being understaffed	Determining task and responsibility sharing
	Inadequate proctoring	
	Failing to meet assessing deadlines	
Washback	Decrease in quality of K12 education	Leaving decision on exam equivalence to universities
	Test-based system (not learning based)	
	Too many exams	Delivering a national four-skills-based exam
	Limited assessment of transferable skills	Ensuring coordination among all units
	Over-reliance on curriculum	Seperating philology students
	Too many MCQ	Conducting a national level needs analysis
	Insufficient impact on production	Transferring 21st-century skills into exams
	Content heavy curriculum	Giving orientation to newcoming students
	Low English proficiency	Choosing up-to-date and authentic exam topics
	Limited of critical thinking	
	Low cognitive-cultural level	
	Internalisation of productive concept	
	No impact of nonclass activities	
Standardisation	No standards	Preparing a common exam calendar
	Constant system change	Establishing HE Language Qualifications

	Instability in the system	Framework
	Ambiguity in the operational system	Working on benchmarking
	Ineffective coordination among TUM	Including EPPs in university visit programmes
	Insufficient coordination among the units	Ensuring mutual recognition
	Quality studies at surface or documentation level	Conducting a national level needs analysis
	Lack of prior documentation	Establishing consortiums or body institutions
	Problems with the previous system	Collaborating with international organizations
Training/Specialization	Lack of pre-service/in-service training	Establishing consortiums or body institutions
	Lack of feedback system	Receiving training from online platforms
	Lack of evaluation system (appraisal or warning)	Giving orientation to newcomers to TUs
	Lack of orientation	Collaborating with international organizations
	Lack of testing knowledge	Establishing consortiums or body institutions
	Need for professional training (item writing, spec)	Giving priority to trainings
	Test security (steel safe, camera, key)	Prioritising TM in assignments
	Validity-reliability matters	Fostering the continuity of professional development
	Item analysis	
	Difficulty in standardization of tests	
Emotional Factors	Demotivation	Maintaining relations of TUM-Instructors-Students
	Burnout	in management level
	Anxiety	Financial incentives by YÖK and TUBITAK
	Top-down approach/assignment	Adding some criteria into Associate Professorship applications
	Excessive testing staff turnover	
	Need for comfortable work environment	Integrating exam preparation into YÖKSİS account
	Unjustified objections	

	Negative attitudes and clashes	Including TM into item writer pool
	Ambiguity in the operational system	
Infrastructure	Technical needs	Conducting a national level needs analysis
	Physical requirement	Determining bottom-up quality parameters
	Lack of specialised human capacity	Giving technical and sanitation staff
	Limited contribution	Integrating EPPs into automation system
	Scarcity in resources and equipment	Leaving a free room for meetings/group work
	Need for support staff	Working from home
Sustainability/Efficiency	Constant system change	Establishing HE Language Qualifications Framework
	Instability in the system	
	Insufficient coordination among the units	Working on benchmarking
	Ineffective coordination among TUMs	Cloud system, locked shelves, archiving, back-up
	Lack of archiving system	Establishing consortiums or body institutions
	Risk management	Developing physical and online archive system
	Lack of prior documentation	Preparing an item pool
Digitalization		Working on digitalisation of exams
	Incompatible automation system	Receiving training from online platforms
	Technology incompetence	Receiving online feedback
	Adaptation to change	Developing online archive system
	Over documentation	Providing licenced programmes needed
	Lack of online exam	Integrating EPPs into automation system
	Need for licenced programmes	Transferring 21st-century skills into exams
		Using electronic open sources

Taking into account all the challenges gained from the interviews with both school directors and testing unit members and the solution suggestions received from the field experts, the next section deals with implications to different stakeholders.

6.2. Implications of the Study

In this part, implications for authority bodies, university/school leaders, and TUMs are presented, respectively.

6.2.1. Implications for authority bodies

Higher education is both a gateway to employment. It contributes to economic development, as well as the professional and academic development of the persons, their social, psychological, and cultural development. It plays a very important role in social development. Thus, it is recommended that action decisions be taken in order to ensure that university students' foreign language skills are at a sufficient level before the employment stop, that is graduation. Since our country has a centralized structure, it is expected that rather than students' individual efforts, authority institutions will be more effective to influence university leaders and students through their firm decisions and actions.

A product of careful planning should be made meaningfully by expanding the decision-making mechanisms to different stakeholders including EPPs. A comprehensive reform in the restructuring of the system of higher education should be undertaken. While this planning is being done, it is suggested that needs analysis studies conducted by MEB should be implemented by other umbrella institutions in a way to include practitioners at universities. In this way, it is believed that decisions and legislative studies can be more realistic and applicable.

The council is recommended to take some steps on necessary legal adaptations to establish the legal ground of the units. In this way, the outcomes of such legal studies could also be applied to other units at these schools. While conducting legal work, the title of the testing unit member and job description should be defined in the legal document. A job description is supposed to play a key role in possible motivation and ensure their well-being and appropriate working conditions if their job descriptions are defined clearly. Otherwise, it seems that having to work in TUs without a legal title

seems to result in demotivation among TUMs. Until a u-turn is made in the national language examination, the universities should be given the freedom to decide on equivalence and recognition of YDS/YÖKDİL.

There are various universities from east to west, from developed to new, with various qualities. One university does not accept a score from another university having a different student type, different institutional cultures, different instructors, different curricula, especially different exam formats, etc. One of the reasons for the difference among universities is the lack of a common umbrella framework in our country; therefore, a framework can be established under the coordination of YÖK, as an authority institution, through workshops and meetings attended by relevant administrators and expert groups from universities. Thus, the council is suggested to establish Higher Education Language Qualifications Framework in which exit levels need to be revised. Regarding the difference, a national language framework has the potential to bring the universities setting the minimum standards and quality.

Financial incentives can be provided by YÖK by giving some credits to the testing members for the preparation of high-stake tests in academic incentive applications. In addition, some criteria can be added to Associate Professorship applications as a testing member may receive some extra points for working in these units. In the past, instructors working in preparatory schools were not allowed to conduct graduate studies. Therefore, the number of people who had a Ph.D. degree was very few. However, as the restrictions were abolished, the number of graduate students increased gradually. One of the goals of each academician with doctorate degree is to pass associate professorship exam. In this case, in the application for associate professorship, giving additional points to those who serve in such offices will increase voluntariness. Likewise, exam preparation could be displayed at YÖKSİS accounts, which is also considered in the reappointment process of academicians. To put it simply, the preparation of large-scale exams can also be added to this system, which will be tangible proof of their products for those who work in TUs.

In educational programmes, language testing courses should be revised to include other topics about assessment and evaluation in the syllabus. That is, the syllabus should be updated by adding recent topics with regard to language assessment to enable prospective students to understand the relationship between language learning and testing rather than only measuring language abilities. In this way, they will be able

to keep up with recent developments. Besides, the course is suggested to embrace integrated testing and assessment. Moreover, a balanced time should be allocated to each topic in the course because some TUMs complained about the greater emphasis put on testing grammar, vocabulary, and receptive skills than productive skills; and thus, there is a need to focus on the ways of assessing especially speaking and writing skills.

National-level incentives have a great place in motivating academicians in Türkiye. Financial incentives to be delivered by TÜBİTAK would foster language instructors to volunteer to be working in these offices. To this aim, a new and special call can be established for language testing.

YÖKAK may include EPPs in university visit programmes. This may ensure that quality parameters about these schools would be determined bottom-up, not top-down. The presence of representatives from EPPs in the university visits of the committees selected by the Quality Board can create a positive basis for the decisions to be taken.

ÖSYM should conduct a new format in language exams and prepare a four-skill-based exam at a national level instead of test-based ones. In other words, a skill-based exam is needed instead of test-oriented exams such as YDS or YÖKDİL to break the impacts of washback of language testing on language learning and teaching. The lack of a four-skill-based exam is supposed to change the implementations of EPPs and their tendencies in assessment.

TUMs can be included in item writer pool by ÖSYM based on their success in exam preparation in their academic accounts. The ones writing exam items for ÖSYM earn extra income. However, it is not clear how and why certain people are selected to write items. To make the system transparent and inclusive, the selection can be done through an application call. Out of the applicants, success points might be used to list the applicants.

There is a need for a guide explaining the key points of developing a language test with integrated skills from the basic level to a deeper level. There are several books on language testing and assessment, yet almost all of them are based on theoretical grounds, which lack practical handbooks showing from zero level. In such a book, promoting practicing skills in language assessment is needed. Otherwise, as in the

present study, the course might remain too theoretical. It is believed that without practice, pre-service teachers cannot develop meaningful assessment knowledge. Therefore, language testing courses should include more examples, exercises, and tasks so that students may bridge the connection between their knowledge and their skills of assessment (Chun, 2010; Sarıyıldız, 2018; Viengsang, 2016).

6.2.2. Implications for EPP Directors and Testing Unit Members

Although it is clearly stated by both the principals and TUMs that it is necessary to carry out legal work, it does not seem possible in the short term to do this work due to the prioritisation of other social and political problems in the country. In this case, it is recommended that university rectors appoint at least one of the philology departments when appointing administrators to these schools. In addition, internal arrangements can be made so that those working in these units can benefit from institutional resources or opportunities in order to make these units more attractive.

Though TUMs are trying to do their best considering the limitations in their offices, there are still some points to be revised to improve testing processes in order to prepare better tests for their future assessment duties. Therefore, in light of the findings, some pedagogical implications can be suggested.

Language assessment differs in terms of its purpose and target group (Jeong, 2013); however, what TUMs should do at first is to do some research through self-study to be aware of what language assessment means (Coombe, Troudi & Al-Hamly, 2012). In the current study, all TUMs were found to be aware of its general meaning; however, they viewed there was no difference between language assessment, language testing, evaluation, and measurement having difficulty in differentiating these terms. In this vein, they need to receive well-organized, intensive training so that their perspectives could be broadened about assessment and other functions of assessment could be comprehended, including how assessment contributes to learning and teaching (DeLuca et al., 2018). Otherwise, their insufficient knowledge and misperceptions about assessment are considered to affect their future assessment performance (Öz & Atay, 2017). During the training, the local needs of Turkish educational system in terms of testing and assessment practices can be discussed and exemplified so they would be

more aware of different concerns and challenges about assessment peculiar to a specific context.

In-service training is not offered as the resources and opportunities of universities and schools are limited. In this case, it is important that those working in these units can benefit from free online training. In addition, they can request help by establishing a WhatsApp group with similar universities in the same province or region and sharing their problems or questions on this channel.

6.3. Suggestions for Further Studies

Higher education is both a gateway to employment. It contributes to economic development, as well as the professional and academic development of the persons, their social, psychological, and cultural development. It plays a very important role in social development.

As mentioned throughout the study, there are some insightful studies focusing on language testing. However, in terms of challenges and solutions in a comprehensive way, the present study can be one of the first studies. Nevertheless, there are still several points to be revealed within the related literature considering the limitations of the present study.

Another suggestion would be to compare the views of students and TUMs in terms of language assessment. In terms of national policies, further research such as this study is needed to be able to compare the results.

There is still a lack of a clear description of what proficiency is in general and how it is assessed. Further investigation is necessary to clarify broadly expressed items in language testing, especially new concepts and implementations since some recent topics are still too general to explain in this study. In addition, it can be investigated to identify individual contexts and factors that affect the choice of how proficiency is described theoretically and how it is determined practically.

The other possibility for further research is to compare testing preferences, and applications of universities in Türkiye. By means of this type of study, the similarities and differences can be better understood. It would be beneficial in order to take some steps towards standardisation by drawing upon comparable research studies.

There is no interview form available focusing on instructors' views on the testing problems and challenges, representing the status of EPPs in Türkiye, which would provide a firm basis for developing solutions.

First of all, the findings of the present study are limited to 50 universities. There are 207 universities in Türkiye most of which have a preparatory school. Further studies may focus on different contexts. The participant group of this study consists of directors and TUMs. Further studies can investigate the subject by including different varieties of stakeholders, including the provincial directorate of YÖK, MEB, TÜBİTAK, YÖKAK, and international stakeholders.

The researcher employed a semi-structured interview form during interviews. However, other data collection techniques such as observation, reflection reports may be used in other studies. A checklist could be developed and applied to universities.

In addition, the present study was conducted by the researcher only. In further studies, working with a team of professionals, field experts, and technical staff in the form of larger projects would be beneficial to gain more insight into the subject.

REFERENCES

- Abduh, M. Y. M. (2021). Full-time online assessment during COVID-19 lockdown: EFL teachers' perceptions. *Asian EFL Journal*, 28(1.1), 26-46.
- Aboudahr, S. M. F. M., & Olowoselu, A. (2018). Analysis of principals decision making styles on teachers performance in selected secondary schools of Gharbia Governorate, Egypt. *Academic Journal of Economic Studies*, 4(4), 91-95.
- Acar-Erdol, T., & Yıldızlı, H. (2018). Classroom assessment practices of teachers in Turkey. *International Journal of Instruction*, 11(3), 587-602.
- Address, P. (2018). In: Personal Communication.
- Akça, S., & Akbulut, M. (2018). Türkiye'deki yağmacı dergiler: Beall listesi üzerine bir araştırma. *Bilgi Dünyası*, 19(2), 255-274.
- Akpınar, K., & Cakildere, B. (2013). Washback effects of high-stakes language tests of Turkey (KPDS and ÜDS) on productive and receptive skills of academic personnel. *Journal of language and linguistic studies*, 9(2), 81-94.
- Akpur, U. (2017). Yıldız Teknik Üniversitesi hazırlık öğretim programının değerlendirilmesi. *Cumhuriyet Uluslararası Eğitim Dergisi*, 6(4), 441-457.
- Aksan, M. M. (2001). *Instructors' perceptions of the content validity of the English language exams at Niğde University* Unpublished Master's Thesis. Ankara: Bilkent University, Institute of Economics and Social Sciences. Retrieved from: <http://www.thesis.bilkent.edu.tr/0001838.pdf>.
- Aktaş, C. K., & Gündoğdu, K. (2020). An extensive evaluation study of the English preparatory curriculum of a foreign language school. *Pegem Journal of Education and Instruction*, 10(1), 169-214.
- Alagozlu, N. (2012). English as a foreign language cul-de-sac in Turkey. *Procedia-Social and Behavioural Sciences*, 47, 1757-1761.
- Albayrak, H., & Şener, T. (2021). The relationship between participation in extracurricular activities and motivation of foreign language learning. *International Journal of Psychology and Educational Studies*, 8(2), 122-132.
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied linguistics*, 14(2), 115-129.
- Allen, D. (2016). Japanese cram schools and entrance exam washback. *The Asian Journal of Applied Linguistics*, 3(1), 54-67.
- Altbach, P. G., & Knight, J. (2007). The internationalisation of higher education: Motivations and realities. *Journal of studies in international education*, 11(3-4), 290-305.
- Altiner, C. (2018). Preparatory school students' english language learning motivation: A sample from Turkey. *Universal Journal of Educational Research*, 6(8), 1729-1737.
- Arik, B. T. (2020). English in Turkey. *World Englishes*, 39(3), 514-527.
- Arik, B. T., & Arik, E. (2014). The role and status of English in Turkish higher education: English is the language of instruction in around 20% of the programs in Turkish universities. *English Today*, 30(4), 5-10.
- Arik, B. T., & Arik, E. (2018). English-medium instruction in Turkish higher education: The current state of English in psychology departments. *Journal of English as an International Language*, 13(1), 20-36.

- Artino Jr, A. R., La Rochelle, J. S., Dezee, K. J., & Gehlbach, H. (2014). Developing questionnaires for educational research: AMEE Guide No. 87. *Medical teacher*, 36(6), 463-474.
- Atar, C., Kır-Cullen, E., & Denkci-Akkaş, F. (2020). An evaluation of the English preparatory program at Istanbul Medeniyet University. *Language Teaching and Educational Research*, 3(1), 94-115.
- Avella, J. R. (2016). Delphi panels: Research design, procedures, advantages, and challenges. *International Journal of Doctoral Studies*, 11, 305-421.
- Aydın, B. (2017). *Türkiye’de hazırlık okullarında yaşanan sorunlara bir çözüm önerisi: Anadolu Üniversitesi örneği*. Ankara: Pegem Akademi.
- Aydın, B., Akay, E., Polat, M., & Geridönmez, S. (2016). Türkiye’deki hazırlık okullarının yeterlik sınavı uygulamaları ve bilgisayarlı dil ölçme fikrine yaklaşımları. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 16(2), 1-20.
- Aygün, Ö. (2017). *A Scale of Turkish Preparatory School University Students’ Demotivational Factors Towards Learning English* Unpublished Masters' Thesis. Eskişehir: Anadolu University, Graduate School of Educational Sciences. Retrieved from: <https://toad.edam.com.tr/sites/default/files/pdf/hazirlik-sinifinda-okuyan-turk-ogrencilerin-ingilizce-ogrenme-motivasyonlarini-azaltan-unsurlar.pdf>.
- Bachman, L., & Palmer, A. (2022). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., Lynch, B. K., & Mason, M. (1995). Investigating variability in tasks and rater judgements in a performance test of foreign language speaking. *Language testing*, 12(2), 238-257.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford: Oxford University Press.
- Bachman, L. F., Palmer, A. S., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.
- Bachmann, L., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.
- Balcı, Ö., Durak-Üğüten, S., & Çolak, F. (2018). Zorunlu İngilizce hazırlık programının değerlendirilmesi: Necmettin Erbakan Üniversitesi yabancı diller yüksekokulu örneği. *Journal of Theoretical Educational Science*, 11(4), 860-893.
- Balcı, Ö., & Üğüten, S. D. (2018). The relationship between EFL learners' language learning strategy use and achievement. *International Education Studies*, 11(2), 1-12.
- Baryshnikova, O., Kostenko, A., & Voskoboynikov, S. (2021). Digital technologies in foreign language learning. In E3S Web of Conferences (Vol. 273). Retrieved from: <https://doi.org/10.1051/e3sconf/202127312144>,
- Bayat, K., & Rezaei, A. (2015). Importance of teachers’ assessment literacy. *International Journal of English Language Education*, 3(1), 139-146.
- Bayraktaroğlu, S. (2014). Yabancı dil eğitimi gerçeği: Yabancı dille eğitim yanılığı. In Sarıçoban Arif & H. Öz (Eds.), *Türkiye’de Yabancı Dil Eğitiminde Eğilim Ne Olmalı?* Ankara: Hacettepe Üniversitesi Yayınları.

- Bayram, I., & Canaran, Ö. (2019). Evaluation of an English preparatory program at a Turkish foundation university. *Journal of language and linguistic studies*, 15(1), 48-69.
- Bechger, T. M., Kuijper, H., & Maris, G. (2009). Standard setting in relation to the Common European Framework of Reference for Languages: The case of the state examination of Dutch as a second language. *Language Assessment Quarterly*, 6(2), 126-150.
- Beckers, D. G., Van der Linden, D., Smulders, P. G., Kompier, M. A., Taris, T. W., & Geurts, S. A. (2008). Voluntary or involuntary? Control over overtime and rewards for overtime in relation to fatigue and work satisfaction. *Work & Stress*, 22(1), 33-50.
- Bellhouse, G. L. (2018). Are teachers given sufficient tools as examiners in high-stakes language testing? A study of the new foreign language speaking component of the French Baccalauréat. *Teacher involvement in high-stakes language testing*, 85-103.
- Benseler, D. P., & Schulz, R. A. (1980). Methodological trends in college foreign language instruction. *The Modern Language Journal*, 64(1), 88-96.
- Berger, J.-L., & Karabenick, S. A. (2016). Construct validity of self-reported metacognitive learning strategies. *Educational Assessment*, 21(1), 19-33.
- Bernard, H. R., Wutich, A., & Ryan, G. W. (2016). *Analysing qualitative data: Systematic approaches*. CA: SAGE publications.
- Berry, V., & O'Sullivan, B. (2016). Assessment literacy for language teachers. *TEA SIG: IATEFL*. <http://tea.iatefl.org/wp-content/uploads/2015/10/Vivien-Berryand-Barry-OSullivan.pdf>.
- Birjandi, P. (2001). The application of reading strategies in Persian and English (A comparative study). *The International Journal of Humanities*, 8(3), 29-36.
- Blalock, M. H. (2014). *The effects of NCLB and recommendations for change: A modified delphi study of principals perceptions*. Knoxville, TN: University of Tennessee.
- Bocanegra-Valle, A. (2014). 'English is my default academic language': Voices from LSP scholars publishing in a multilingual journal. *Journal of English for Academic Purposes*, 13, 65-77.
- Bourdeaux, C., & Chikoto, G. (2008). Legislative influences on performance management reform. *Public Administration Review*, 68(2), 253-265.
- Brantley, D. K. (2007). *Instructional assessment of English language learners in the K-8 classroom*. Boston: Allyn and Bacon.
- Brindley, G. (2003). Classroom-based assessment. In *Practical English language teaching* (pp. 309-328). New York: McGraw-Hill.
- British Council. (2014). *Türkiye'deki devlet okullarında İngilizce dilinin öğretimine ilişkin ulusal ihtiyaç analizi*. Ankara: Yorum Basın Yayın Sanayi.
- British Council. (2015). *The state of English in Higher Education in Turkey: A Baseline Study*. Ankara: TEPAV.
- Brown, G. (2001). *Assessment: A guide for lecturers* (Vol. 3). York, LTSN Generic Centre.
- Brown, G. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318.
- Brown, J. (1996). 1996: Testing in language programs. Upper Saddle River, NJ: Prentice Hall Regents.

- Brown, S. (2005). Assessment for learning. *Learning and teaching in higher education*, 1(1), 81-89.
- Bryant, P. C., & Allen, D. G. (2013). Compensation, benefits and employee turnover: HR strategies for retaining top talent. *Compensation & Benefits Review*, 45(3), 171-175.
- Bülbül, T. (2003). Ankara Üniversitesi Eğitim Bilimleri Fakültesinde görev yapan öğretim üyelerinin lisansüstü öğretime öğrenci seçme sürecine ilişkin görüşleri. *Ankara University Journal of Faculty of Educational Sciences (JFES)*, 36(1), 167-174.
- Burgess, S., & Sievertsen, H. H. (2020). Schools, skills, and learning: The impact of COVID-19 on education. *VoxEu. org*, 1(2). Retrieved from: <https://voxeu.org/article/impact-covid-19-education> (accessed on 28 May 2020).
- Burton, D. M., & Bartlett, S. (2020). Introduction to education studies. *Introduction to Education Studies*, 1-464.
- Butler-Kisber, L. (2018). Constant comparison inquiry. *Qualitative inquiry. London: SAGE Publications*, 41-59.
- Büyükahıska, D. (2020). English language testing and evaluation course through the lens of EFL student teachers. *Rumeli Dil ve Edebiyat Araştırmaları Dergisi*(19), 581-596.
- Büyükkarcı, K. (2016). Identifying the areas for English language teacher development: A study of assessment literacy. *Pegem Journal of Education and Instruction*, 6(3), 333-346.
- Canagarajah, A. S. (2005). *Reclaiming the local in language policy and practice*. Routledge.
- Carr, D. (2005). *Making sense of education: An introduction to the philosophy and theory of education and teaching*. Routledge.
- Carr, N. T. (2011). *Designing and analysing language tests*. Oxford: Oxford University Press.
- Carr, S. M. (2011). NCLEX-RN pass rate peril: One school's journey through curriculum revision, standardized testing, and attitudinal change. *Nursing Education Perspectives*, 32(6), 384-388.
- Çatı, K., Kethüda, Ö., & Bilgin, Y. (2016). Positioning strategies of universities: An investigation on universities in İstanbul. *Education and Science/Eğitim ve Bilim*, 41(185), 219-234.
- Cephe, P., & Toprak, E. (2014). The common European framework of reference for languages: Insights for language testing. *Journal of language and linguistic studies*, 10(1), 79-88.
- Cesur, M. O. (2011). Can language learning strategies predict Turkish university prep class students' achievement in reading comprehension? *Procedia-Social and Behavioural Sciences*, 15, 1920-1924.
- Çetintaş, B. (2010). Türkiye'de yabancı dil eğitim ve öğretiminin sürekliliği. *Journal of language and linguistic studies*, 6(1), 65-74.
- Chan, K. L. R. (2020). Washback in education: A critical review and its implications for language teachers. *Journal of Foreign Language Education and Technology*, 5(1), 108-124.
- Charmaz, K., & Belgrave, L. (2012). Qualitative interviewing and grounded theory analysis. *The SAGE handbook of interview research: The complexity of the craft*, 2, 347-365.

- Cheng, H.-F., & Dörnyei, Z. (2007). The use of motivational strategies in language instruction: The case of EFL teaching in Taiwan. *International Journal of Innovation in Language Learning and Teaching*, 1(1), 153-174.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study* (Vol. 21). Cambridge: Cambridge University Press.
- Cheng, L., & Curtis, A. (2004). *Washback or backwash: A review of the impact of testing on teaching and learning*. Routledge.
- Christoforidou, M., & Kyriakides, L. (2021). Developing teacher assessment skills: The impact of the dynamic approach to teacher professional development. *Studies in Educational Evaluation*, 40, 1–11.
- Chun, M. (2010). Taking teaching to (performance) task: Linking pedagogical and assessment practices. *Change: The magazine of higher learning*, 42(2), 22-29.
- Cocchetta, F. (2018). Developing university students' multimodal communicative competence: Field research into multimodal text studies in English. *System*, 77, 19-27.
- Çokluk, Ö., Yılmaz, K., & Oğuz, E. (2011). Nitel bir görüşme yöntemi: Odak grup görüşmesi. *Kuramsal Eğitimbilim*, 4(1), 95-107.
- Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges, and recommendations. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyneff (Eds.), *The Cambridge guide to second language assessment* (pp. 20-29). Cambridge: Cambridge University Press.
- Coskun, A. (2013). An investigation of the effectiveness of the modular general English language teaching preparatory program at a Turkish university. *South African Journal of Education*, 33(3), 1-18.
- Council of Europe (CoE). (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe (CoE). (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment-companion volume*. Strasbourg: Council of Europe Publishing. www.coe.int/lang-cefr
- Creswell, J. W. (2013). Steps in conducting a scholarly mixed methods study. 48. Retrieved from: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1047&context=db_erspeaker.
- Creswell, J. W., & Creswell, J. D. (2005). Mixed methods research: Developments, debates, and dilemmas. *Research in organisations: Foundations and methods of inquiry*, 2, 315-326.
- Dadey, N., Lyons, S., & DePascale, C. (2018). The comparability of scores from different digital devices: A literature review and synthesis with recommendations for practice. *Applied Measurement in Education*, 31(1), 30-50.
- Davis, A. (1995). Criterion-referenced assessment and the development of knowledge and understanding. *Journal of Philosophy of Education*, 29(1), 3-21.
- Dearden, J. (2015). *English as a medium of instruction-a growing global phenomenon*. London: British Council.
- Dearden, J., Macaro, E., & Akincioglu, M. (2016). English medium instruction in universities: A collaborative experiment in Turkey. *Studies in English Language Teaching*, 4(1), 3–34.

- DeLuca, C., Chapman-Chin, A. E., LaPointe-McEwan, D., & Klinger, D. A. (2018). Student perspectives on assessment for learning. *The Curriculum Journal*, 29(1), 77-94.
- Demir, H., Kapukaya, K., & Özfıdan, B. (2015). Yabancı diller yüksek okullarında görev yapan ingilizce okutmanlarının sorunları. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 12(30), 113-138.
- Demir, Y. (2012). *A comparative analysis on soft students' perceptions regarding nest and non-nest effects on motivation and attitudes* (Unpublished Master Dissertation). Selçuk University, Institute of Educational Sciences, Konya, Turkey.
- Demiray-Akbulut, F. (2016). ESP needs analysis of university preparatory school students: Learning-centred approach. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 13(36), 181-202.
- Demirtaşlı-Çıktıkçı, N. (2002). Lisansüstü eğitim programlarına girişte lisansüstü eğitimi giriş sınavı (LES) sonucunun ve diğer ölçütlerin kullanımına ilişkin bir tarama. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 35(1-2), 61-70.
- Denzin, N. K., & Lincoln, Y. S. (2011). *The Sage handbook of qualitative research*. Thousand Oaks, CA: Sage.
- Denzin, N. K., Lincoln, Y. S., & Giardina, M. D. (2006). Disciplining qualitative research. *International journal of qualitative studies in education*, 19(6), 769-782.
- Doğan, D. (2013). Problems of newly-founded universities and solution offers. *Journal of Higher Education and Science*, 3(2), 108-116.
- Doğan, N., Koyuncu, İ., Gökdemir, P., & Kahveci, M. (2016). Öğrencilerin eğitim bilimleri enstitüsü lisansüstü programlarına kabul durumlarının yordanması. *The Journal of Educational Research*, 2(2), 114-131.
- Doğan, S. (2018). 1992 yılında kurulan türk üniversiteleri rektörlerinin 2023 hedeflerinin incelenmesi. *Yükseköğretim ve Bilim Dergisi*(2), 336-347.
- Drew, G. (2010). Issues and challenges in higher education leadership: Engaging for change. *The Australian educational researcher*, 37(3), 57-76.
- Ecclestone, K. (2005). *Learning autonomy in post-16 education: The policy and practice of formative assessment*. Routledge.
- EF English Proficiency Index. (2019). *EF English proficiency index: a ranking of 100 countries and regions by English skills*. <https://www.ef.com/wwen/epi/>
- EF English Proficiency Index. (2022). *EF English proficiency index: a ranking of 111 countries and regions by English skills*. <https://www.ef.com>
- Eignor, D. R. (2001). Standards for the development and use of tests: The standards for educational and psychological testing. *European Journal of Psychological Assessment*, 17(3), 157-163.
- Ekoç, A. (2021a). Mobile Language Learning Applications from the Perspectives of Adult Language Learners in Turkey. *Shanlax International Journal of Education*, 9(4), 259-264.
- Ekoç, A. (2021b). Mapping university preparatory school A1 level students' high school English learning experience. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, 22, 865-882.
- Enisa, M., & Uygun, S. (2014). Evaluation of a language preparatory program: A case study. *ELT Research Journal*, 3(4), 201-221.

- Erarslan, A. (2019). Progressive vs modular system in preparatory school English language teaching program: A case of system change at a state university in Turkey. *Journal of language and linguistic studies*, 15(1), 83-97.
- Erarslan, A. (2021). English language teaching and learning during COVID-19: A global perspective on the first year. *Journal of Educational Technology and Online Learning*, 4(2), 349-367.
- Erdem, A. R. (2013). Bilgi toplumunda üniversitenin değişen rolleri ve görevleri. *Yükseköğretim Dergisi*, 3(2), 109-120.
- Erdem, C., & Atar, C. (2018). Yükseköğretim düzeyindeki zorunlu İngilizce derslerinin öğretim görevlisi görüşlerine göre değerlendirilmesi. ICPESS-VENICE. 28-30 June 2018. Venice-Italy.
- Erdoğan, A. (2010). *Yükseköğretimde yeniden yapılanma: 66 soruda Bologna Süreci uygulamaları*. Yükseköğretim Kurulu.
- Erdoğan, G., & Mede, E. (2021). The evaluation of an English preparatory program using CIPP model and exploring A1 level students' motivational beliefs. *Journal of Education and Educational Development*, 8(1), 53-76.
- Erdoğan, T. (2018). The investigation of self-regulation and language learning strategies. *Universal Journal of Educational Research*, 6(7), 1477-1485.
- Erfiani, N. N., & Ngadiso, S. (2018). The washback effects of testing practices on English language teaching. *Indonesian Journal of English Teaching*, 7(2), 103-109.
- Erkuş, H., Tolga, O., & Sever, S. (2020). Üniversite öğrencilerinin bölüm tercihlerini etkileyen faktörler üzerine bir araştırma. *Dicle Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 10(20), 261-275.
- Ersoy, N. Ş., & Yapıcıoğlu, D. K. (2015). İsteğe bağlı İngilizce hazırlık programının öğrenci ve okutman görüşlerine göre değerlendirilmesi. *Eğitimde Nitel Araştırmalar Dergisi*, 3(3), 7-43.
- Ertürk, N. O. (2015). Testing your tests: Reliability issues of academic English exams. *International Journal of Psychology and Educational Studies*, 2(2), 47-52.
- Ferguson, G., Pérez-Llantada, C., & Plo, R. (2011). English as an international language of scientific publication: A study of attitudes. *World Englishes*, 30(1), 41-59.
- Fişne, M. (2016). Yeni Türk dış politikasını etkileyen etik dış politika yaklaşımının incelenmesi. *Afyon Kocatepe Üniversitesi Sosyal Bilimler Dergisi*, 18(2), 153-169.
- Flowerdew, J. (1999). Problems in writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing*, 8(3), 243-264.
- Flowerdew, J. (2008). Scholarly writers who use English as an Additional Language: What can Goffman's "Stigma" tell us? *Journal of English for Academic Purposes*, 7(2), 77-86.
- Foster, D. (2013). Security issues in technology-based testing. In J. A. Wollack & J. J. Fremer (Eds.), *Handbook of test security* (pp. 261-283). New York, NY: Routledge.
- Foster, D., & Layman, H. (2013). *Online proctoring systems compared*. Retrieved on 5 February 2018. <https://ivetriedthat.com/wp-content/uploads/2014/07/Caveon-Test-Security.pdf>.
- Franco-Santos, M., & Doherty, N. (2017). Performance management and well-being: a close look at the changing nature of the UK higher education workplace. *The International Journal of Human Resource Management*, 28(16), 2319-2350.

- Fulcher, G. (2004). Deluded by artifices? The common European framework and harmonisation. *Language Assessment Quarterly: An International Journal*, 1(4), 253-266.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132.
- Fulcher, G. (2013). Philosophy and language testing. *The companion to language assessment*, 3, 1431-1451.
- Fulcher, G. (2014). *Testing second language speaking*. New York, NY: Routledge.
- Gebril, A. (2017). Language teachers' conceptions of assessment: An Egyptian perspective. *Teacher Development*, 21(1), 81-100.
- Genesee, F., Upshur, J. A., & Richards, J. C. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.
- Giraldo-Aristizábal, F. (2018). A diagnostic study on teachers' beliefs and practices in foreign language assessment. *Íkala, revista de lenguaje y cultura*, 23(1), 25-44.
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile Issues in Teachers Professional Development*, 20(1), 179-195.
- Gökçe, A., & Batman, O. (2015). Turizm ön lisans programlarında meslek amaçlı İngilizce (EOP) öğretiminin öğrenci görüşlerine dayalı ihtiyaç değerlendirmesi. *Paradoks Ekonomi Sosyoloji ve Politika Dergisi*, 211-264.
- Gömleksiz, M. N., & Erten, P. (2010). Mesleki ve teknik ortaöğretimde uygulanan modüler öğretim programının etkililiğinin değerlendirilmesi (Elazığ ili örneği). *Education Sciences*, 5(3), 1037-1055.
- Gonzales, E. (2011). An EFL placement test and its use in a private high school. *Open Distance Learning Master of Arts in Teaching English as a Foreign/Second Language (MA TEFL/TESL)*. University of Birmingham.
- Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, 33(2), 14-18.
- Granot, E., Brashear, T. G., & Motta, P. C. (2012). A structural guide to in-depth interviewing in business and industrial marketing research. *Journal of Business & Industrial Marketing*, 27(7), 547-553.
- Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education*, 14(1), 75-97.
- Green, A. (2013). Washback in language assessment. *International Journal of English Studies*, 13(2), 39-51.
- Green, A. B., & Weir, C. J. (2004). Can placement tests inform instructional decisions? *Language testing*, 21(4), 467-494.
- Guilloteaux, M. J., & Dörnyei, Z. (2008). Motivating language learners: A classroom-oriented investigation of the effects of motivational strategies on student motivation. *TESOL quarterly*, 42(1), 55-77.
- Gümüş, S., & Gülmez, D. (2020). Yükseköğretimde kurumsal yönetim: Akademik birimlerin yönetim süreci ve karşılaşılan sorunlar. *Yükseköğretim Dergisi*, 10(1), 73-84.
- Günay, D., & Günay, A. (2017). Türkiye'de yükseköğretimin tarihsel gelişimi ve mevcut durumu. *Yükseköğretim Dergisi*, 7(3), 156-178.
- Gündoğdu, K., Çelik, B., Hancı-Yanar, B., Yolcu, O., & Ceylan, V. K. (2016). Türkiye'de yükseköğretimde program geliştirme bağlamında uluslararasılaşma. *Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi Dergisi*(33), 35-48.
- Güven, A., & Gough, C. (2018). In: Personal Communication.

- Hakim, B. (2015). English language teachers' ideology of ELT assessment literacy. *International Journal of Education and Literacy Studies*, 3(4), 42-48.
- Hanafin, S. (2004). Review of literature on the Delphi Technique. *Dublin: National Children's Office*, 1-51.
- Harrison, D. A., Newman, D. A., & Roth, P. L. (2006). How important are job attitudes? Meta-analytic comparisons of integrative behavioural outcomes and time sequences. *Academy of Management journal*, 49(2), 305-325.
- Hatipoğlu, Ç. (2010). Summative evaluation of an English language testing and evaluation course for future English language teachers in Turkey. *English Language Teacher Education and Development (ELTED)*, 13, 40-51.
- Hatipoğlu, Ç. (2015). English language testing and evaluation (ELTE) training in Turkey: Expectations and needs of pre-service English language teachers. *ELT Research Journal*, 4(2), 111-128.
- Hatipoğlu, Ç., & Erçetin, G. (2016). Türkiye'de yabancı dilde ölçme ve değerlendirme eğitiminin geçmişi ve bugünü. Proceedings of the 3. Ulusal Yabancı Dil Eğitimi Çalıştayı: Türkiye'deki Yabancı Dil Eğitimi Üzerine Görüş ve Düşünceler, Istanbul: Boğaziçi University Press.
- Haznedar, B. (2010). Türkiye'de yabancı dil eğitimi: Reformlar, yönelimler ve öğretmenlerimiz. International Conference on New Trends in Education and Their Implications. (pp. 747-755).
- Hidri, S. (2018). Introduction: State of the art of assessing second language abilities. In *Revisiting the assessment of second language abilities: From theory to practice* (pp. 1-19). Cham: Springer.
- Hildén, R., & Fröjdendahl, B. (2018). The dawn of assessment literacy: Exploring the conceptions of Finnish student teachers in foreign languages. *Apples-Journal of Applied Language Studies*, 12(1), 1-24.
- Huang, H., Liu, L., Yang, S., Cui, X., Zhang, J., & Wu, H. (2019). Effects of job conditions, occupational stress, and emotional intelligence on chronic fatigue among Chinese nurses: a cross-sectional study. *Psychology research and behaviour management*, 12, 351.
- Huang, L., Kubelec, S., Keng, N., & Hsu, L. (2018). Evaluating CEFR rater performance through the analysis of spoken learner corpora. *Language Testing in Asia*, 8(1), 1-17.
- Hughes, A. (2003). *Testing for language teachers*. Ernst Klett Sprachen.
- Hulstijn, J. H., Alderson, J. C., & Schoonen, R. (2010). Developmental stages in second-language acquisition and levels of second-language proficiency: Are there links between them. *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*, 11-20.
- Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, 31, 58-69.
- İlkuçar, M. (2011). Sınav gözetmenlik çizelgeleme probleminin optimizasyonu ve bir uygulama yazılımı. *XIII. Akademik Bilişim Konferansı*, 2-4.
- İnal, B., & Aksoy, E. (2014). Çankaya Üniversitesi hazırlık sınıfı İngilizce öğretim programının değerlendirilmesi. *Journal of Research in Education and Teaching*, 3(3), 119-134.
- Iqbal, H. S., Akhtar, M. M. S., & Saleem, M. (2020). A study of decision-making styles of academic managers in public sector universities of the Punjab. *Bulletin of Education and Research*, 42(2), 181-196.

- Irgatoğlu, A. (2017). *An Analysis into Teacher Self-Development Models in One-Year Preparatory Schools of Private Universities in Ankara*. Unpublished Master's Thesis. Ankara: Gazi Üniversitesi, Eğitim Bilimleri.
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language testing*, 30(3), 345-362.
- Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language testing*, 27(4), 555-584.
- Johnson, J. M., & Rowlands, T. (2012). The interpersonal dynamics of in-depth interviewing. *The SAGE handbook of interview research: The complexity of the craft*, 99-113.
- Kachru, B. B. (1999). The other side of English and the 1990s. In B. B. Kachru (Ed.), *The other tongue: English across cultures*. University of Illinois Press.
- Karakaş, A. (2015). Orientations towards English among English-medium instruction students. *Online Submission*, 2(1), 1-38.
- Karakaş, A. (2017). The forgotten voices in higher education: students' satisfaction with English-medium instruction. *Online Submission*, 12(1), 1-14.
- Karacı-Aktas, C., & Gündoğdu, K. (2020). An extensive evaluation study of the English preparatory curriculum of a foreign language school. *Pegem Journal of Education and Instruction*, 10(1), 169-214.
- Kartal, G., & Basol, H. Ç. (2019). Generic teacher competencies and the English language teacher education program in Turkey. *Novitas-ROYAL*, 13(2), 133-154.
- Kayahan-Karakul, A., & Karakütük, K. (2014). Views of academics on student selection criteria for graduate education. *Ankara University Journal of Faculty of Educational Sciences (JFES)*, 47(2), 179-200.
- Keser, A. D. (2018). *An investigation on the exit criteria of English language preparatory programs of Turkish universities: a Delphi method analysis* [Anadolu University (Turkey)].
- Keser, A. D., & Durmuşoğlu-Köse, G. (2019). Determining exit criteria for English Language proficiency in preparatory programs at Turkish universities. *The Online Journal of Quality in Higher Education*, 6(2), 45.
- Kılınç, A. Ç., Çepni, O., Kılcan, B., & Palaz, T. (2017). Öğretim üyelerinin görüşlerine göre Türkiye de yeni kurulan üniversitelerin sorunları: Nitel bir araştırma. *Yükseköğretim ve Bilim Dergisi*(2), 282-293.
- Kinman, G., & Johnson, S. (2019). Special section on well-being in academic employees. *International Journal of Stress Management*, 26(2), 159.
- Kır, E., & Sülü, A. (2014). Language teachers' views on CEFR. *International Online Journal of Education and Teaching (IOJET)*, 1(5), 358-364.
- Kırkgöz, Y. (2007). English language teaching in Turkey: Policy changes and their implementations. *RELC journal*, 38(2), 216-228.
- Kırkgöz, Y. (2014). Students' perceptions of English language versus Turkish language used as the medium of instruction in higher education in Turkey. *Electronic Turkish Studies*, 9(12).
- Kırkgöz, Y. (2017). English education policy in Turkey. In *English language education policy in the Middle East and North Africa* (pp. 235-256). Springer.
- Klassen, A. C., Creswell, J., Plano Clark, V. L., Smith, K. C., & Meissner, H. I. (2012). Best practices in mixed methods for quality of life research. *Quality of life Research*, 21(3), 377-380.

- Kocman, A., & Weber, G. (2018). Job satisfaction, quality of work life and work motivation in employees with intellectual disability: A systematic review. *Journal of Applied Research in Intellectual Disabilities*, 31(1), 1-22.
- Koğar, H., & Sayın, A. (2014). Yüksek lisans programına giriste öğrencilerin sınıflandırılma doğruluğunun İncelenmesi: Gazi Üniversitesi örneği. *Ankara University, Journal of Faculty of Educational Sciences*, 47(1), 171-186.
- Köksal, D., & Cesur, K. (2012). Students and instructors' perceptions of objective tests used to assess language performance at university level 1. *Eğitimde Kuram ve Uygulama Journal of Theory and Practice in Education*, 8(1), 41-53.
- Kolb, S. M. (2012). Grounded theory and the constant comparative method: Valid research strategies for educators. *Journal of emerging trends in educational research and policy studies*, 3(1), 83-86.
- Korkmaz, H. E., & Erden, M. (2014). A Delphi study: The characteristics of democratic schools. *The Journal of Educational Research*, 107(5), 365-373.
- Kostecká, Y., Kostecký, T., Vodičková, K., & Jančářík, A. (2015). Linguistic integration of middle school immigrant children in Czechia. *AUC GEOGRAPHICA*, 50(2), 181-192.
- Krosnick, J. A. (2018). Questionnaire design. In *The Palgrave handbook of survey research* (pp. 439-455). Springer.
- Küçüköğlü, M., & Taş, H. Y. (2022). Kamu harcamaları ve mali aldanma ilişkisi. *Hak İş Uluslararası Emek ve Toplum Dergisi*, 11(31), 263-276.
- Külekcî, E. (2016). A concise analysis of the foreign language examination (YDS) in Turkey and its possible washback effects. *International Online Journal of Education and Teaching*, 3(4), 303-315.
- Kumar, S., & Gautam, N. (2018). Decision making styles among professor in central university of Bihar-An empirical study of predictors. *International Journal of Law and Society*, 1(2), 84-91.
- Kunnan, A. J. (2004). Regarding language assessment. *Language Assessment Quarterly: An International Journal*, 1(1), 1-4.
- LaFlair, G. T., Langenfeld, T., Baig, B., Horie, A. K., Attali, Y., & von Davier, A. A. (2022). Digital-first assessments: A security framework. *Journal of Computer Assisted Learning*, 38(4), 1077-1086.
- Lee, E.-S., Park, T.-Y., & Koo, B. (2015). Identifying organisational identification as a basis for attitudes and behaviours: A meta-analytic review. *Psychological bulletin*, 141(5), 1049.
- Leech, N. L., & Onwuegbuzie, A. J. (2011). Beyond constant comparison qualitative data analysis: Using NVivo. *School Psychology Quarterly*, 26(1), 70.
- Leong, P., Joseph, S. R., & Boulay, R. (2010). Applying constant comparative and discourse analyses to virtual worlds research. *Journal For Virtual Worlds Research*, 3(1), 3-26.
- Lillis, T. M., & Curry, M. J. (2010). *Academic writing in global context*. Routledge London.
- Linn, R. L. (2006). The standards for educational and psychological testing: Guidance in test development. In T. M. Haladyna (Ed.), *Handbook of Test Development* (pp. 27-38). Mahwah, NJ: Lawrence Erlbaum.
- Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2004). A call for standardisation in content analysis reliability. *Human Communication Research*, 30(3), 434.
- Lum, G. (2012). Two concepts of assessment. *Journal of Philosophy of Education*, 46(4), 589-602.

- Luxia, Q. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language testing*, 22(2), 142-173.
- MacIntyre, P. D., Gregersen, T., & Mercer, S. (2020). Language teachers' coping strategies during the Covid-19 conversion to online teaching: Correlations with stress, wellbeing and negative emotions. *System*, 94, 1-13. <https://doi.org/https://doi.org/10.1016/j.system.2020.102352>.
- Madsen, H. S. (1983). *Techniques in Testing*. ERIC.
- Martinez-Duboc, A. P. (2009). Language assessment and the new literacy studies. *Lenguaje*, 37(1), 159-178.
- McGunagle, D., & Zizka, L. (2020). Employability skills for 21st-century STEM students: the employers' perspective. *Higher Education, Skills and Work-Based Learning*, 10, 591-606.
- McNamara, T. (2000). *Language testing*. Oxford University Press.
- McNamara, T., Knoch, U., Fan, J., & Rossner, R. (2019). *Fairness, justice & language assessment*. Oxford University Press.
- Mede, E., & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Dil Dergisi*, 168(1), 1-5.
- Merç, A. (2015). Using technology in the classroom: A study with Turkish pre-service EFL teachers. *Turkish Online Journal of Educational Technology-TOJET*, 14(2), 229-240.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage Publications.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2018). *Qualitative data analysis: A methods sourcebook*. Sage Publications.
- Mimirinis, M. (2019). Qualitative differences in academics' conceptions of e-assessment. *Assessment and Evaluation in Higher Education*, 44(2), 233-248.
- Mondada, L. (2007). Commentary: Transcript variations and the indexicality of transcribing practices. *Discourse studies*, 9(6), 809-821.
- Moser, K. M., Wei, T., & Brenner, D. (2021). Remote teaching during COVID-19: Implications from a national survey of language educators. *System*, 97, 102431. <https://doi.org/https://doi.org/10.1016/j.system.2020.102431>.
- Mustajab, D., Azies, B., Irawan, A., Rasyid, A., Akbar, M. A., & Hamid, M. A. (2020). COVID-19 pandemic: What are the challenges and opportunities for e-Leadership? *Fiscaoekonomia*, 4(2), 483-497.
- Mutlu, G. (2018). A program evaluation study of the main course at a preparatory program: A case study. *Turkish Online Journal of Qualitative Inquiry*, 9(3), 202-239.
- Mutlu, G. (2020). Türk akademisyenlerin yağmacı dergilere ilişkin görüşleri: Bir nitel betimsel çalışma. *Eğitimde Nitel Araştırmalar Dergisi*, 8(1), 107-134.
- Ng, T. W., Yam, K. C., & Aguinis, H. (2019). Employee perceptions of corporate social responsibility: Effects on pride, embeddedness, and turnover. *Personnel Psychology*, 72(1), 107-137.
- Nworie, J. (2011). Using the Delphi technique in educational technology research. *TechTrends*, 55(5), 24-30.
- O'Hara, L., De Souza, L. H., & Ide, L. (2000). A Delphi study of self-care in a community population of people with multiple sclerosis. *Clinical rehabilitation*, 14(1), 62-71.

- Ogunode, N. J., & Okwelogu, I. S. (2022). Investigation into causes of inadequate academic staff in public universities in north-central Nigeria. *Miasto Przyszłości*, 25, 308-313.
- Oğuz, E., & Karakaya, İ. (2009). Lisansüstü eğitim programlarının değerlendirilmesi (ondokuz mayıs üniversitesi örneği). 4. *Lisansüstü Eğitim Sempozyumu Bildiriler*. Ayşe Çakır İlhan, Müge Artar, H. Hüseyin Aksoy (Ed). Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Yayınları, 193-204.
- Oktay, A. (2015). Foreign language teaching: A problem in Turkish education. *Procedia-Social and Behavioural Sciences*, 174, 584-593.
- Ölmezer-Öztürk, E., & Aydın, B. (2019). Investigating language assessment knowledge of EFL teachers. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 34(3), 602-620.
- Ölmezler-Öztürk, E. (2019). *Developing and validating language assessment knowledge scale (LAKS) and exploring the assessment knowledge of EFL teachers* Anadolu University (Turkey)].
- Öner, G., & Mede, E. (2015). Evaluation of A1 level program at an English preparatory school in a Turkish university: a case study. *ELT Research Journal*, 4(3), 204-226.
- Öz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, 6(1), 25-44.
- Özer, S. (2019). Turizm işletmeciliği bölümü öğrencilerinin ingilizce ve mesleki ingilizce kavramlarına ilişkin metaforik algıları. *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi*, 10(1), 364-380.
- Özkanal, Ü., & Hakan, A. G. (2010). Effectiveness of university English preparatory programs: Eskisehir Osmangazi University foreign languages department English preparatory program. *Journal of Language Teaching and Research*, 1(3), 295-305.
- Öztürk, M. (2015). Yükseköğretimde modüler yabancı dil eğitim programı: Hacettepe Üniversitesi örneği. *International Journal of Curriculum and Instructional Studies*, 4(8), 113-128.
- Öztürk, N. (2019). Yükseköğretim hazırlık sınıfı dil eğitmeni istihdamına yönelik eleştirel bir söylem. *Yükseköğretim ve Bilim Dergisi*(3), 375-383.
- Öztürk, Z. (2015). Opinions of university preparatory school students towards the effects of learning English on their career. *Procedia-Social and Behavioural Sciences*, 199, 785-792.
- Özüdoğru, F. (2017). Evaluation of the voluntary English preparatory program at a Turkish state university. *Journal of International Social Research*, 10(48), 501-509.
- Özusu, N. (2018). *An evaluation of the performance assessment component of an EFL preparatory school program* Yeditepe Üniversitesi, Eğitim Bilimleri Enstitüsü].
- Paker, T. (2014). Yabancı dil eğitimi ve öğretiminde programların tüm süreçlerinin denetimi ve bireysel yönlendirmenin rolü. In S. Arif & H. Öz (Eds.), *Türkiye'de Yabancı Dil Eğitiminde Eğilim Ne Olmalı?* Ankara: Hacettepe Üniversitesi Yayınları.
- Papadima-Sophocleous, S. (2022). Teacher education for language assessment and testing: Postgraduate program evaluation from its students' perspective. *Colombian Applied Linguistics Journal*, 24(1), 67-88.
- Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice*. Sage publications.

- Pearson, W. S. (2021). Policies on minimum English language requirements in UK higher education, 1989-2021. *Journal of Further and Higher Education*, 45(9), 1240-1252.
- Perneger, T. V., Courvoisier, D. S., Hudelson, P. M., & Gayet-Ageron, A. (2015). Sample size for pre-tests of questionnaires. *Quality of life Research*, 24(1), 147-151.
- Pitoyo, M. D., Sumardi, S., & Asib, A. (2020). Gamification-based assessment: The washback effect of Quizizz on students' learning in higher education. *International Journal of Language Education*, 4(2), 1-10.
- Polster, C. (2016). Vicious circle: Academic insecurity and privatisation in western universities. In *Handbook of the sociology of higher education* (pp. 94-105). New York, NY: Routledge.
- Powell, C. (2003). The Delphi technique: myths and realities. *Journal of advanced nursing*, 41(4), 376-382.
- Raišienė, A. G., Rapuano, V., Varkulevičiūtė, K., & Stachová, K. (2020). Working from home—Who is happy? A survey of Lithuania's employees during the COVID-19 quarantine period. *Sustainability*, 12(13), 5332.
- Rashidi, N., & Javanmardi, F. (2011). The IELTS preparation washback on learning and teaching outcomes. *Cross-Cultural Communication*, 7(3), 132-144.
- Raworth, K., Sweetman, C., Narayan, S., Rowlands, J., & Hopkins, A. (2012). *Conducting semi-structured Interviews*. Oxfam.
- Read, J. (2015). *Assessing English proficiency for university study*. Springer.
- Robins, C. S., & Eisen, K. (2017). Strategies for the effective use of NVivo in a large-scale study: Qualitative analysis and the repeal of Don't Ask, Don't Tell. *Qualitative Inquiry*, 23(10), 768-778.
- Rogler, D. (2014). Assessment literacy: Building a base for better teaching and learning. *English Teaching Forum*, 52(3), 2-13.
- Roulston, K., & Choi, M. (2018). Qualitative interviews. *The SAGE handbook of qualitative data collection*, 233-249.
- Şahinel, S. (1997). *Ankara Üniversitesi hazırlık sınıflarında uygulanmakta olan İngilizce sınavlarına ilişkin okutmanlarının görüşleri* Unpublished MA Thesis. Ankara: Hacettepe University, Institute of Social Sciences.
- Saldana, J. (2011). *Fundamentals of qualitative research*. Oxford University Press.
- Salehi, H., & Yunus, M. M. (2012). The washback effect of the Iranian universities entrance exam: Teachers' insights. *GEMA Online Journal of Language Studies*, 12(2), 609-628.
- Saraç, Y. (2019). *Araştırma ve aday araştırma üniversiteleri 2017 ve 2018 yılları performans izleme ve değerlendirme sonuçları toplantısı*. Yükseköğretim Kurulu.
- Saraç, Y. (2020). *YÖK 100/2000 Doktora Projesi*. Yükseköğretim Kurulu.
- Sarıçoban, G., & Sarıçoban, A. (2012). Atatürk and the history of foreign language education in Turkey. *Journal of language and linguistic studies*, 8(1), 24-49.
- Sarıyıldız, G. (2018). A study into language assessment literacy of preservice English as a foreign language teachers in Turkish context.
- Savić, D. (2020). COVID-19 and work from home: Digital transformation of the workforce. *Grey Journal (TGJ)*, 16(2), 101-104.
- Savignon, S. J. (2005). Communicative language teaching: Strategies and goals. In *Handbook of research in second language teaching and learning* (pp. 659-676). Routledge.

- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language testing*, 30(3), 309-327.
- Schlechter, A., Thompson, N. C., & Bussin, M. (2015). Attractiveness of non-financial rewards for prospective knowledge workers: An experimental investigation. *Employee Relations*, 37(3), 273-295.
- Seçilmiş, C., & Ünlüöner, K. (2009). Modüler öğretim sisteminin getirdiği uygulamaların değerlendirilmesi: Anadolu otelcilik ve turizm meslek liselerinde bir alan araştırması. *İşletme Araştırmaları Dergisi*, 1(2), 3-18.
- Seidman, G. (2013). Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personality and individual differences*, 54(3), 402-407.
- Selvi, B. (2021). English as the language of science: The role of English in the academic world. *Akademik Hassasiyetler*, 8(16), 1-21.
- Semiz, Ö., & Odabaş, K. (2016). Turkish EFL teachers' familiarity of and perceived needs for language testing and assessment literacy. *Proceedings of LILA*, 16, 66-72.
- Şenay, H. H., Şengül, M., & Seggie, F. N. (2020). Türkiye'de Yükseköğretim Çalışmaları: Eğilimler ve Öneriler. *Üniversite Araştırmaları Dergisi*, 3(1), 1-13.
- Senyshyn, R. M. (2002). *Cross-cultural competencies in international management curricula: A Delphi study of faculty perspectives*. The University of Tennessee.
- Serinikli, N. (2021). COVID-19 salgın sürecinde örgütsel değişim: Uzaktan/evden çalışma modeli. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 31(1), 277-288.
- Serpil, H. (2000). *An assessment of the content validity of the midterm achievement tests administered at Anadolu University Foreign Languages Department* [Bilkent University].
- Setiyana, R. (2016). Analysis of summative tests for English. *English Education Journal*, 7(4), 433-447.
- Sevimel-Şahin, A., & Subaşı, G. (2019). An overview of language assessment literacy research within English language education context. *Journal of Theoretical Educational Science*, 12(4), 1340-1364.
- Shariff, N. (2015). Utilizing the Delphi survey approach: A review. *J Nurs Care*, 4(3), 246.
- Sheehan, S., & Munro, S. (2017). *Assessment: attitudes, practices, needs* (Vol. 17). ELT Research Papers, 17.08. London, UK: British Council. <https://www.teachingenglish.org.uk>
- Shohamy, E. (1997). Testing methods, testing consequences: Are they ethical? Are they fair? *Language testing*, 14(3), 340-349.
- Shohamy, E. (2001). *The power of tests: a critical perspective of the uses of language tests*. Longman.
- Simons, H. (2009). *Case study research in practice*. London: SAGE.
- Şişman, E. P., & Büyükkarcı, K. (2019). A review of foreign language teachers' assessment literacy. *Sakarya University Journal of Education*, 9(3), 628-650.
- Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education: Research*, 6(1), 1-21.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language teaching research*, 9(1), 5-29.

- Stabler-Havener, M. L. (2018). Defining, conceptualizing, problematizing, and assessing language teacher assessment literacy. *Studies in Applied Linguistics and TESOL*, 18(1).
- Stansfield, C., & Winke, P. (2008). Testing aptitude for second language learning. *Encyclopaedia of language and education, 2nd Edition: Language Testing and assessment*, 7, 81-94.
- Sultana, N. (2018). A brief review of washback studies in the south Asian countries. *The Educational Review, USA*, 2(9), 468-474.
- Suna, Y., & Durmuşçelebi, M. (2013). Türkiye’de yabancı dil öğrenme-öğretme problemine ilişkin yapılan çalışmaların derlemesi. *OPUS Uluslararası Toplum Araştırmaları Dergisi*, 3(5), 31-48.
- Syamananda, P. (2017). Factors Affecting EFL Teachers' Motivation in Thai University: A Case Study of EFL Teachers at Tertiary Level. *LEARN Journal: Language Education and Acquisition Research Network*, 10(2), 120-131.
- Szmerdt-Chandler, D. (2011). Assessing oral production. *Issues in Promoting Multilingualism*, 257-286.
- Tannenbaum, R. J., & Wylie, E. C. (2008). Linking English-language test scores onto the common European framework of reference: An application of standard-setting methodology. *ETS Research Report Series, 2008*(1), 1-75.
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154-155.
- Techera, U., Hallowell, M., Stambaugh, N., & Littlejohn, R. (2016). Causes and consequences of occupational fatigue. *Journal of occupational and environmental medicine*, 58(10), 961-973.
- Tekin, M. (2015). Evaluation of a preparatory school program at a public university in Turkey. *Journal of International Social Research*, 8(36), 718-733.
- Tekneci, P. D. (2016). Türk yükseköğretim sisteminin son onyılda dönüştürümü. *Yükseköğretim ve Bilim Dergisi*, 6(3), 277-287.
- Téllez, K., & Mosqueda, E. (2015). Developing teachers’ knowledge and skills at the intersection of English language learners and language assessment. *Review of Research in Education*, 39(1), 87-121.
- Tercan, G. (2018). Evaluating the modular system of preparatory class program. *ELT Research Journal*, 7(1), 2-23.
- Thibault-Landry, A., Gagne, M., Forest, J., Guerrero, S., Seguin, M., & Papachristopoulos, K. (2017). The relation between financial incentives, motivation, and performance. *Journal of personnel Psychology*, 16, 61–76.
- Thomas, G. (2011). A typology for the case study in social science following a review of definition, discourse, and structure. *Qualitative Inquiry*, 17(6), 511-521.
- Timothy, A., & Ryan, K. (2008). Job Satisfaction: Subjective well-being at work. In *The science of subjective well-being* (pp. 393-413). The Guilford Press New York.
- Todd, R. W. (2020). Teachers’ perceptions of the shift from the classroom to online teaching. *International Journal of TESOL Studies*, 2(2), 4-16.
- Toprak, M., Kolat, D., Şengül, M., & Erdoğan, A. (2019a). Türk Yükseköğreniminin Görünümü: Betimsel Bir Analiz. *Üniversite Araştırmaları Dergisi*, 2(2), 113-130.
- Toprak, M., Kolat, D., Şengül, M., & Erdoğan, A. (2019b). Türk yükseköğreniminin ISCED geniş ve dar alan sınıflaması. *Üniversite Araştırmaları Dergisi*, 2(3), 138-154.

- Tosuncuoğlu, İ. (2018). English Language Teaching as lingua franca. *Karaelmas Eğitim Bilimleri Dergisi*, 6(2), 326-331.
- 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.
- Tuzcu-Eken, D. (2021). English education in Turkey: Why are we 'very low level'? *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*(23), 804-822.
- Ülker, N. (2017). Küreselleşen dünyada Türkiye'deki Yükseköğretim kurumlarında İngilizce öğretimi: Sorunlar ve çözüm önerileri. *İstanbul Teknik Üniversitesi Vakıf Dergisi*, 78, 19-22.
- Ulum, Ö. G. (2015). The situational analysis of the physical state of an ELT department. *International Journal of Social Sciences and Education Research*, 2(3), 949-959.
- Ulum, Ö. G., & Uzun, K. (2020). Critical perspective of English teaching and learning in Turkey. *International Journal of Evaluation and Research in Education*, 9(2), 456-460.
- Unal, M., & Ilhan, E. (2017). A case study on the problems and suggestions in foreign language teaching and learning at higher education. *Journal of Education and Training Studies*, 5(6), 64-72.
- Ünal, S., Sarı, İ., & Gürol, M. (2017). Evaluation of the English language teaching curriculum. *Pegem Atıf İndeksi*, 609-618.
- UNESCO. (2020). <https://www.unesco.org/en>
- UNESCO. (2021). <https://www.unesco.org/en>
- Urbina-Garcia, A. (2020). What do we know about university academics' mental health? A systematic literature review. *Stress and Health*, 36(5), 563-585.
- Üstunluoglu, E., Zazaoglu, K. F. A., Keskin, M. N., Saraykoylu, B., & Akdogan, G. (2012). Developing a CEFR based curriculum: A case study. *Online Submission*, 5(1), 115-128.
- Viengsang, R. (2016). Exploring pre-service English teachers' language assessment literacy. *Modern Journal of Language Teaching Methods*, 6(5), 432-442.
- Vogt, K., Tsagari, D., Csépes, I., Green, A., & Sifakis, N. (2020). Linking learners' perspectives on language assessment practices to teachers' assessment literacy enhancement (TALE): Insights from four European countries. *Language Assessment Quarterly*, 17(4), 410-433.
- von Davier, A. A. (2017). Computational psychometrics in support of collaborative educational assessments. *Journal of Educational Measurement*, 54(1), 3-11.
- Wächter, B., & Maiworm, F. (2014). *English-taught programmes in European higher education: The state of play in 2014*. Lemmens Medien GmbH.
- Wei, W. (2017). A critical review of washback studies: Hypothesis and evidence. In R. Al-Mahrooqi, C. Coombe, F. Al-Maamari, & V. Thakur (Eds.), *Revisiting EFL Assessment: Critical Perspective* (pp. 49-67). Cham: Springer.
- Weir, C. J. (2005a). Language testing and validation. *Hampshire: Palgrave MacMillan*.
- Weir, C. J. (2005b). Limitations of the Common European Framework for developing comparable examinations and tests. *Language testing*, 22(3), 281-300.
- Widarko, A., & Anwarodin, M. K. (2022). Work motivation and organisational culture on work performance: Organisational citizenship behaviour (OCB) as mediating variable. *Golden Ratio of Human Resource Management*, 2(2), 123-138.

- Wind, S. A., & Peterson, M. E. (2018). A systematic review of methods for evaluating rating quality in language assessment. *Language testing*, 35(2), 161-192.
- World Health Organisation (WHO). (2020). <https://www.who.int/>
- Woznyj, H. M., Banks, G. C., Whelpley, C. E., Batchelor, J. H., & Bosco, F. A. (2022). Job attitudes: A meta-analytic review and an agenda for future research. *Journal of Organisational Behaviour*, 43(5), 946-964.
- Yaman, İ. (2018). Türkiye’de İngilizce öğrenmek: zorluklar ve fırsatlar. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*(11), 161-175.
- Yastıbaş, A. E., & Takkaç, M. (2018). Understanding language assessment literacy: Developing language assessments. *Journal of language and linguistic studies*, 14(1), 178-193.
- Yıldırım, A., & Şimşek, H. (2006). *Sosyal bilimlerde nitel araştırma yöntemleri*. Ankara: Seçkin Yayınları.
- Yükseköğretim Kurulu (YÖK). (2016). *Yükseköğretim kurumlarında yabancı dil öğretimi ve yabancı dille öğretim yapılmasında uyulacak esaslara ilişkin yönetmelik*. Ankara: YÖK.
- Yükseköğretim Kurulu (YÖK). (2019). “Yağmacı” dergi yayınları akademik yükseltmelerde kullanılamayacak
- Yükseköğretim Kurulu (YÖK). (2021). *Türkiye Yükseköğretim Sistemi-2021*. Ankara: YÖK.
- Zhao, Z. (2013). An overview of studies on diagnostic testing and its implications for the development of diagnostic speaking test. *International journal of English linguistics*, 3(1), 41.
- http-1:** <https://istatistik.yok.gov.tr/>. (Retrieved on 15.12.2021)
- http-2:** <https://uis.unesco.org/>. (Retrieved on 04.12.2022).
- http-3:** <https://www.iienstitu.com/>. (Retrieved on 15.12.2021).
- http-4:** <https://www.mevzuat.gov.tr>. (Retrieved on 15.12.2021).
- http-5:** <https://yokak.gov.tr/>. (Retrieved on 15.12.2021).
- http-6:** <https://studyportals.com/>. (Retrieved on 15.12.2021).

APPENDICES

APPENDIX 1. Ethics Committee Approval

APPENDIX 2. Consent Form

APPENDIX 3. Interview Form for Testing Members

APPENDIX 4. Interview Form for Directors/Vice Directors

APPENDIX 5. Full list of challenges in Phase I

APPENDIX 6. Full list of solutions in Phase 2

APPENDIX 7. Sample Transcription

APPENDIX 8. Curriculum Vitae

APPENDIX 1. Ethics Committee Approval

Evrak Kayıt Tarihi: [REDACTED]

Protokol: [REDACTED]

Tarih: 24.06.2019



ANADOLU ÜNİVERSİTESİ

SOSYAL VE BEŞERÎ BİLİMLER BİLİMSEL ARAŞTIRMA VE YAYIN ETİĞİ KURULU

KARAR BELGESİ

ÇALIŞMANIN TÜRÜ:	Doktora Tez Çalışması
KONU:	Eğitim Bilimleri
BAŞLIK:	Türk Yükseköğretim Kurumlarındaki İngilizce Hazırlık Okullarında Ölçme: Problemler, Çözüm Önerileri Assessment in the Preparatory Language Schools in Turkish Higher Education Institutions: Problems, Solutions
PROJE/TEZ YÜRÜTÜCÜSÜ:	Doç. Dr. Ali MERÇ
TEZ YAZARI:	Mehmet ŞENGÜL
ALT KOMİSYON GÖRÜŞÜ:	-
KARAR:	Olumlu

[REDACTED]

APPENDIX 2. Consent Form

Gönüllü Katılım Formu

Üniversite Hazırlık Okullarında yapılan sınavlar üzerine planlanan bu çalışma, Anadolu Üniversitesi İngilizce Dili Öğretmenliği Doktora öğrencisi Mehmet Şengül tarafından yürütülmektedir. Bu çalışmada amacım, üniversite hazırlık okullarında yapılan sınavlarda yaşanan sorunların tespiti ve bu konuda getirilebilecek çözüm yollarını araştırmaktır. Gerek akademik ortamlarda gerekse ilgili toplantılarda hazırlık okullarında yapılan sınavlarda birtakım sorunların olduğu ve bu konuda çözüm önerilerinin bulunması gerektiği dile getirilmektedir. Ancak bu konuda üniversite bazında ve mikro düzeyde bir veya birkaç husus üzerine araştırma yapılmış olsa da büyük resmi görmek adına bütüncül bir çalışma bulunmamaktadır.

Üniversite hazırlık okullarında yapılan sınavlarda yaşanan sorunların tespitine yönelik görüş ve düşüncelerinizi öğrenmek istiyorum. Bu amaçla size ilgili sorular yöneltilecek ve görüşmemizi onay verdiğiniz ölçüde ses kaydı olarak daha sonra incelenmesi amacıyla kaydedeceğim. Görüşmemiz iki kısımdan oluşmaktadır. İlk kısım, sizin demografik özellikleriniz üzerine soruları kapsamakta olup, ikinci kısımda ise hazırlık okullarında yapılan sınav süreçlerine yönelik sorulardan oluşmaktadır. Çalışmaya önceden belirlenen takvim doğrultusunda katılmanız istenecektir. Çalışma süresince ve sonrasında kimlik bilgileriniz proje dışındaki hiç kimseyle izniniz dışında paylaşılmayacaktır (Ancak üniversite ismi veya katılımcı ismi hiçbir yerde belirtilmeyecektir). Bu çalışma kapsamında elde edilecek olan bilimsel bilgiler sadece araştırmacılar tarafından yapılan bilimsel yayınlarda, sunumlarda ve eğitim amaçlı çevrimiçi bir ortamda katılımcı ismi verilmeden paylaşılacaktır. Görüşme, genel olarak kişisel rahatsızlık verecek soruları içermemektedir. Ancak, katılım sırasında sorulardan ya da herhangi başka bir nedenden ötürü kendinizi rahatsız hissederseniz cevaplama işini yarıda bırakmakta serbestsiniz. Bu çalışmaya katıldığınız için şimdiden teşekkür ederim. Çalışma hakkında daha fazla bilgi almak ve yanıtlanmasını istediğiniz sorularınız için araştırmayı yürüten araştırmacı ile iletişim kurabilirsiniz.

Bu çalışmaya tamamen gönüllü olarak katılıyorum ve istediğim zaman yarıda kesip çıkabileceğimi biliyorum. Bu proje kapsamında gerçekleştirilen görüşmelerde yer alacağımı biliyorum. Verdiğim bilgilerin bilimsel amaçlı yayınlarda kullanılmasını kabul ediyorum. Proje süresince onay verdiğim ölçüde ses kaydı alınacağını biliyorum. Toplanan verilerin bilimsel makaleler, yayınlar, akademik sunumlar ve çevrimiçi bir eğitim ortamı dışında kesinlikle kullanılmayacağını biliyorum.

Çalışmaya katılmak istiyorum.

Tarih:

Ad Soyadı:

İmza:

APPENDIX 3. Interview Form for Testing Members

SINAV BİRİMİ GÖRÜŞME FORMU

Anadolu Üniversitesi İngilizce Dili Öğretmenliği Doktora öğrencisiyim. Üniversite Hazırlık Okullarında yapılan sınavlar üzerine bir araştırma yapıyorum. Amacım, üniversite hazırlık okullarında yapılan sınavlarda yaşanan sorunların tespiti ve bu konuda getirilebilecek çözüm yollarını araştırmaktır.

Üniversite hazırlık okullarında yapılan sınavlarda yaşanan sorunların tespitine yönelik düşüncelerinizi ve beklentilerinizi öğrenmek istiyorum. Görüşmeye katıldığınız için şimdiden teşekkür ederim.

GÖRÜŞME SORULARI

1. Sınav biriminde görev yaparken yönetimle herhangi bir zorlukla karşılaşmıyorsunuz?
2. Soru hazırlayan kişiler olarak akademik, mesleki ve kişisel açıdan ne tür sorunlar yaşamaktasınız?
3. Sınav süreçlerine yansıyan sistemik problemler nelerdir?
4. Sınavlar hazırlanırken ne tür problemler yaşamaktasınız?
5. Sınavların uygulanması sırasında ne tür problemler yaşamaktasınız?
6. Sınavların değerlendirilmesinde ne tür problemler yaşamaktasınız?
7. Sınavların değerlendirilmesi sonrasında ne tür problemler yaşamaktasınız?
8. Yukarıda sorduğum hususlar dışında belirtmek istediğiniz başka bir husus var mıdır?

APPENDIX 4. Interview Form for Directors/Vice Directors

YÖNETİCİ GÖRÜŞME FORMU

Anadolu Üniversitesi İngilizce Dili Öğretmenliği Doktora öğrencisiyim. Üniversite Hazırlık Okullarında yapılan sınavlar üzerine bir araştırma yapıyorum. Amacım, üniversite hazırlık okullarında yapılan sınavlarda yaşanan sorunların tespiti ve bu konuda getirilebilecek çözüm yollarını araştırmaktır.

Üniversite hazırlık okullarında yapılan sınavlarda yaşanan sorunların tespitine yönelik düşüncelerinizi ve beklentilerinizi öğrenmek istiyorum. Görüşmeye katıldığınız için şimdiden teşekkür ederim.

GÖRÜŞME SORULARI

1. Sınav birimi grubunu yönetirken herhangi bir zorlukla karşılaşılıyor musunuz?
2. Soru hazırlayan kişiler akademik, mesleki ve kişisel açıdan ne tür sorunlar yaşamaktadır?
3. Sınav süreçlerine yansıyan sistemik problemler nelerdir?
4. Sınav soruları hazırlanırken ne tür problemler yaşamaktasınız?
5. Sınavların uygulanması sırasında ne tür problemler yaşamaktasınız?
6. Sınavların değerlendirilmesinde ne tür problemler yaşamaktasınız?
7. Sınavların değerlendirilmesi sonrasında ne tür problemler yaşamaktasınız?
8. Yukarıda sorduğum hususlar dışında belirtmek istediğiniz başka bir husus var mıdır?

APPENDIX 5. Full list of challenges in Phase I

Category	Sub-Cat	Codes	Freq (n)	Per (%)
Systemic	National	1-Legal ground	341	16.77
		2-Uncompensated extra work	302	14.86
		3-Lack of job description	269	13.23
		4-Washback effect of YDS/YÖKDİL	231	11.36
		5-No standards	169	8.31
		6-Decrease in quality of K12 education	150	7.38
		7-Test-based system	142	6.98
		8-Over-increase in quotas	135	6.64
		9-Lack of online exam	96	4.72
		10-Need to operate as a separate unit	78	3.84
		11-Pressure for the success of some students	52	2.56
		12-Quality studies at surface or documentation level	41	2.02
		13-Extra exam for late accepted students	27	1.33
	Sub-category 1: National - Total / % sss		2,033	34.75
	Institutional	1-Lack of pre-service/in-service training	332	16.62
		2-Constant system change	273	13.66
		3-Poor physical conditions	254	12.71
		4-Limited decision making	226	11.31
		5-Problems with equipment	205	10.26
		6-Limited financial contribution/support	179	8.96
		7-Insufficient coordination among the units	157	7.86
		8-Incompatible automation system	132	6.61
		9-Lack of feedback system	107	5.36
		10-Lack of evaluation system	64	3.20
		11-Term based system	43	2.15

		12-Limited assessment of transferable skills	26	1.30
	Sub-category 2: Institutional - Total / % sss		1,998	34.15
	Unit	1-Excessive workload	359	19.72
		2-Being understaffed	311	17.09
		3-Excessive testing staff turnover	263	14.45
		4-Need for support staff	230	12.64
		5-Lack of orientation	202	11.10
		6-Ambiguity in the operational system	176	9.67
		7-Need for comfortable work environment	124	6.81
		8-Ineffective coordination among TUMs	89	4.89
9-Problems with the previous system		66	3.63	
Sub-category 3: Unit - Total / % sss		1,820	31.10	
Systemic - Grand Total / %		Category 1: 5,851	33.91	
Stakeholder	Admin	1-Fait accompli	296	24.24
		2-Administrators change	264	21.62
		3-Administrators from outside the ELT field	216	17.69
		4-Inattentive administrators	178	14.58
		5-Lack of testing knowledge	172	14.09
		6-Negative attitudes	95	7.78
	Sub-category 1: Administration - Total / % sss		1,221	25.42
TU	1-Demotivation	404	23.68	

	Ms	2-Need for professional training	363	21.28
		3-Secretarial paperwork	311	18.23
		4-Stress	250	14.65
		5-Need to stay in office	135	7.91
		6-Need for isolation	97	5.69
		7-Teaching and test preparation at the same time	80	4.69
		8-Adaptation to change	66	3.87
	Sub-category 2: Testing Member - Total / % sss		1,706	35.51
	Inst	1-Tense reactions	186	23.91
		2-Distraction	142	18.25
		3-Irresponsibility	123	15.81
		4-Lack of emphaty	88	11.31
		5-Low work discipline	66	8.48
		6-Technology incompetence	65	8.35
		7-Over-reliance on curriculum	58	7.46
		8-Too many instructors	50	6.43
	Sub-category 3: Lecturer - Total / % sss		778	16.19
	Stud	1-Too many students	170	15.47
		2-Low English proficiency	165	15.01
		3-Gap (Difference) between levels	159	14.47
		4-False level placement	132	12.01
		5-Unwillingness	127	11.56
		6-Reactions to tests	111	10.10
		7-Lack of of critical thinking	85	7.73
		8-Internalisation of productive concepts	75	6.83
		9-Lack of information in optic forms	43	3.91
		10-Low cognitive-cultural level	32	2.91

	Sub-category 4: Student - Total / % sss		1,099	22.88
Stakeholder - Grand Total / %			Category 2: 4,804	27.84
Testing Process	Prep	1-Approximation to levels	131	13.86
		2-Difficulty in item preparation	128	13.55
		3-Scarcity in resources	124	13.12
		4-Lack of reference framework	117	12.38
		5-Insufficient proofreading	106	11.22
		6-Need for native speaker	97	10.26
		7-Validity matters	82	8.68
		8-Alignment matters	73	7.73
		9-Lack of item pool	52	5.50
		10-Lack of topic variety	35	3.70
	Sub-category 1: Preparation - Total / % sss		945	23.93
	Imp	1-Large scale exams	153	21.89
		2-Technical issues	145	20.74
		3-Inadequate proctoring	139	19.89
		4-Inadequate instructions	102	14.59
		5-Latecomers to exams	76	10.87
		6-No recording in speaking exams	51	7.30
		7-Reading aloud listening parts	33	4.72
	Sub-category 2: Implementation - Total / % sss		699	17.70
	Evaluation	1-Reliability issues	224	19.48
		2-Difficulty in standardization	187	16.26

		3-Lack of well-developed rubric	161	14.00
		4-Failing to meet assessing deadlines	143	12.43
		5-No blind marking	135	11.74
		6-Unclear answer key	114	9.91
		7-Borderline scores	92	8.00
		8-Assessment of papers outside the school	63	5.48
		9-Complicated score calculation	31	2.70
	Sub-category 3: Evaluation - Total / % sss		1,150	29.12
	Post -Ev	1-Over documentation	254	21.99
		2-Unjustified objections to scores	189	16.36
		3-Lack of archiving system	167	14.46
		4-Lack of statistical analysis	152	13.16
		5-Move up/down adaptations	134	11.60
		6-Requirement for a fast announcement	104	9.01
		7-Need for licenced programmes	87	7.53
		8-Lack of prior documentation	68	5.89
	Sub-category 4: Post-evaluation - Total / % sss		1,155	29.25
Category 3: Testing		3,949	22.89	
Procedures - Grand Total / %				
Other	Exam	1-Too many exams	337	28.11
		2-Test security	215	17.93
		3-Risk management	185	15.43
		4-Non-prep exams	180	15.01
		5-Too many multiple-choice questions	131	10.93
		6-Nonappealing tasks	89	7.42
		7-Unbalance in score weighting	62	5.17

	Sub-category 1: Examination - Total / % sss		1,199	45.26
	Educ	1-Insufficient impact on production	145	40.06
		2-Content-heavy curriculum	122	33.70
		3-No impact of non-class activities	95	26.24
	Sub-category 2: Educational - Total / % sss		362	13.67
	Time	1-Time restrictions	425	39.06
		2-Waste of time	378	34.74
		3-Having to skip some processes	285	26.20
	Sub-category 3: Time - Total / % sss		1,088	41.07
Category 4:		2,649	15.35	
Other - Grand Total / %				
Challanges Grand Total / %		17,253	100.00	

APPENDIX 6. Full list of solutions in Phase 2

Category	Sub-Category	Codes	Frequency (n)	Percentage (%)
Category 1: National Solutions				
National	Legal studies	1-Establishing the legal ground	29	11.03
		2-Defining TM title and job description	25	9.51
	Comparable skill-based exam	3-Delivering a national four-skills-based exam	23	8.75
		4-Leaving decision on exam equivalence to universities	21	7.98
		5-Establishing HE Language Qualifications Framework	20	7.60
		6-Ensuring mutual recognition	20	7.60
	Incentives	7-Financial incentives by YÖK and TUBITAK	19	7.22
		8-Adding some criteria into Associate Professorship applications	18	6.84
		9-Integrating exam preparation into YÖKSİS account	17	6.46
		10-Including TM into item writer pool	17	6.46
		11-Including EPPs in university visit programs	16	6.08

	Buttom-up needs analysis	12-Conducting a national level needs analysis	13	4.94
		13-Determining bottom-up quality parameters	10	3.80
		14-Regarding the capacity of EPPs for quotas	9	3.42
		15-Establishing consortiums or body institutions	6	2.28
Total			263	37.09
Category 2: Institutional Solutions				
Institu tional	Training on digitalisati on	1-Working on digitalisation of exams	19	8.09
		2-Giving priority to trainings	18	7.66
		3-Prioritising TM in assignments	17	7.23
		4-Collaborating with international organizations	17	7.23
	Physical and technical needs	5-Improving physical conditions	16	6.81
		6-Giving technical and sanitation staff	16	6.81
		7-Providing licenced programs needed	15	6.38
		8-Integrating EPPs into automation system	14	5.96
	Structural	9-Dividing preparatory education into three	13	5.5

	priorities	terms		3
		10-Adopting level-based system	13	5.5 3
		11-Appointing the EPP director as vice rector or advisor	12	5.1 1
		12-Requiring service in one of the units at EPPs	12	5.1 1
		13-Taking comments of TUs on key issues	11	4.6 8
	Coordination and orientation	14-Ensuring coordination among all units	10	4.2 6
		15-Providing orientation to students	8	3.4 0
		16-Fostering the continuity of professional development	7	2.9 8
		17-Preparing a common exam calendar	6	2.5 5
	Firm authority	18-Maintaining relations of TUM-Instructors-Students in management level	5	2.1 3
		19-Authorising TU for complaints/objections	4	1.7 0
		20-Taking sanction decisions	2	0.8 5
Total			235	33. 15
Category 3: Unit-level Solutions				
Unit-	Systemic	1-Transferring 21st-century skills into exams	21	9.9

level	changes			5
		2-Using electronic open sources	20	9.48
		3-Seperating philology students	19	9.00
		4-Setting up a bottom-up circulation system	18	8.53
	Language testing issues	5-Preparing a table of specifications	17	8.06
		6-Working on validity and reliability matters	17	8.06
		7-Writing well-organized instructions	16	7.58
		8-Choosing up-to-date and authentic exam topics	16	7.58
		9-Determining task and responsibility sharing	15	7.11
		10-Preparing a clear workflow	15	7.11
		11-Setting up a dynamic proofreading cycle	14	6.64
	Specialization	12-Giving orientation to newcomers	13	6.16
		13-Receiving training from online platforms	12	5.69
		14-Having membership to related associations	12	5.69
	Evaluation issues	15-Working on standardization	11	5.21
		16-Determine reliability processes	10	4.7

				4
		17-Preparing and updating well-developed rubrics	10	4.74
	Test security	18-Working on risk management	9	4.27
		19-Collaborating with TUs at other universities on exam security	8	3.79
		20-Leaving a free room for meetings/group work	8	3.79
	Post-evaluation issues	21-Receiving online feedback	7	3.32
		22-Employing statistical analysis	6	2.84
		23-Working on benchmarking	5	2.37
		24-Developing physical and online archive system	4	1.90
		25-Preparing an item pool	3	1.42
Total		211	29.76	
Grand Total		709		

APPENDIX 7. Sample Transcription

GÖRÜŞME

1. Sınav biriminde görev yaparken yönetimle herhangi bir zorlukla karşılaşılıyor musunuz?

T1: Bazı yöneticilerin gözünde hocalarımız müfredat öğretmeniyiz.

T2: Bizim tam olarak ne zaman başladığımız bile belli değil. Başlangıç tarihi olur ya bir bölümün mesela, “artık başlarız” gibi bir şey oluyor. Ders bitince buraya geldik gibi bir şey oluyor.

T3: Ben nisanda benim elimden ders alınarak ben buraya geldim mesela. Testing ofiste çok ihtiyaç var diye buraya geldim. Derslerim bölüştürüldü benim diğer hocalar arasında buraya geldim bir anda.

T1: Yasal zemini olmadığı için böyle.

T3: Aynen. Ama ben testingde çalışırken bir yanda o dersin şeylerini de devam etmek zorunda kaldım. Çocukların coursebook projeleri, writingleri başka hoca bilmez ki onları. Ama hem burada çalıştım hem onlara baktım bir yandan da.

Int: Doğru, ama o da hem sizi yoruyor.

T3: Ekstra bir külfet oldu aslında.

T4: Ben aslında Yıldız’da yeni başladım, başlar başlamaz da testinge geçtim gibi bir şey oldu.

Int: Ek olarak söylemek isterseniz, ara ara söyleyebilirsiniz.

T2: Ben ilk testinge başladığımda soruyorum soruyorum hocam yöneticiler bazı sorulara cevap veremiyor item analysis nedir diyorum, önemliyse yapın diyor. Tabii hoca nerden bilsin eğer testing biriminde çalışmadıysa.

T4: Öyle, cevap veremediğini ben de farkettim.

Int: Bu yüzden ben de yöneticileri ayrı sorular, yönetsel düzeyde sorular sorayım.

T3: Yöneticiler çok uğraşmak istemiyor, ilgilenmek istemiyor, zaten arkadaşlar yapıyor diyor. Sen gece mi çalışmışsın, evde mi çalışmışsın çok bir anlamı olmuyor.

T1: Şuanda bizde öyle olmuyor pek, müdür beyle müdür yardımcımız çok olayın içine müdahiller.

T2: Bundan önce değildi çünkü bizim önceki müdürlerimiz mühendisti, hepsi.

T4: O da çok büyük bir sorun aslında. Şimdi ELT tabanlı biri olduğu için daha işin içine dahil oldu tabii.

T1: Ama devlet üniversitelerinde özellikle çoğu üniversitede öyle hocam maalesef onun çok büyük sıkıntısı var işte ilgilenmek istese bile.

T3: Title yok alanla alakalı olduğu için.

T1: Hoca profesör, müdür. Profesör olması yetmiyor. Sen bir şey söylüyorsun, hoca tamam yapalım diyor ama hoca ne olduğunu bilmiyor.

T2: Hepimiz 5’er saat, dönem başında daha fazla derse giriyoruz.

T3: Bu da bizi yoruyor. Girmek istemediğimizde müdür bey zorluyor bu göreviniz diyor.

T4: fiziksel koşullar açısından Teknik şeyler yine. Gerekli destek verilmiyor teknik ekipmanlarla ilgili.

T1: not edin lütfen, hocaların teknoloji özürlü olması da bu sürece çok büyük etki ediyor. Ortaçağ skolastik dönemindeki gibi yeni bir şey yapmaya korkuyoruz, teknolojik anlamda.

T2: USB ye geçtiğimizde evet bayağı bir direnen oldu, “ay ben nasıl yapacağım” şeklinde, ama öğrendiler en nihayetinde. 4 ya da 5. sınavın sonunda öğrendiler.

T4: Ee hocam daha kolay CD’den.

T3: Evet aslında anlatamadık. Ona alışmış ya.

T4: Biz bilgisayar bile kullanmayı zor ikna ettik.

2. Soru hazırlayan kişiler olarak kişisel açıdan ne tür sorunlar yaşamaktasınız?

T1: Bence şu maddi işle, sekreteryaya işiyle akademik işini birbirinden ayırsalar iyi olur. Okulda bize bunlar öğretilmedi çünkü optik saymak gibi şeyler. Onun ayrılması, akademik title ise akademik title olsun.

T2: Zaman konusunda da belki bu yüzden bu işler okulda yapılması gerektiği için Nurgül hocanında söylediği gibi eve kalıyor, akademik işler daha çok evde yapmamız gerekiyor.

T1: Dikkat gerektiren akademik işler aslında.

T3: Burada daha fiziksel çalışıyormuşuz da diğer işler eve kalıyormuş gibi.

Int: Bu arada hocam eklemek istediğiniz bir şey olursa illa her soruya cevap vereceksiniz diye değil ama eklemek istediğiniz bir şey olursa ekleyebilirsiniz dinlerim ben soruyla ilgili.

T4: O söyledikleri çoğuna hepimiz katılıyoruz zaten.

T1: Kişi eksikliği bence.

T2: Bazen teknik problemler oluyor.

T4: evet, teknik problemlerle ilgili sorunlar yaşıyoruz.

T3: Bastıklarının yarısı şey çıktı okunmuyor, baştakiler değil arada okunmayanlar var zaten öyle olduğu için baştakileri de tek tek çekiyoruz. Hangisi okunuyor, hangisi okunmuyor. Aradan hiç beklemediğimiz bir kağıttan çıkıyor.

T4: Çok büyük bir sorun.

Int: Bir şey söyleyecektiniz hocam.

T1: Bizim normalde derse giren hocalarla bu kadar iletişimimiz olması gerekmiyor, biz biraz daha soyutlanmış olarak çalışmalıyız aslında. Bu sınav süreçlerinde falan çok fazla iç içe kalmamız gerekebiliyor ve o bizim çalışmalarımızı sekteye uğratabiliyor.

T2: Yani dikkatimizi dağıtıyor.

T4: Biraz daha soyut kalabiliriz.

T3: Bir de hocam bu şey şeklinde oluyor maalesef hocaların gereksiz soruları olabiliyor ya da clash şeklinde bazen olabiliyor.

T1: Evet kağıt al ver, gidip gelmeler olsun. Mümkün olduğunca buradaki yeni sistemde şu karşınızdaki gördüğünüz dosyalama sistemi falan iyileştirilmeye çalışılıyor ama tamamiyle ayrılmış bir halde değil.

T2: Öğrenci sayısının çok fazla olması bence büyük sıkıntı, hem sekreteryaya işini çok fazla artırıyor yani 3000 küsür öğrenci var, çok öğrenci var. Orada bir standardizasyon tutturmak da çok zor oluyor. Curriculum olarak, biz curriculum ile bağımlı çalışıyoruz ya, biz mesela bir yere kadar öngörüp sınav hazırlıyoruz, birden curriculumdan bir cevap geliyor, bazı sınıflar oraya kadar gelememiş, çünkü çok fazla sınıf var. Her öğrenciye aynı **pacing de** ders verilemiyor, profiller farklı sınıflarda kiminin algılaması daha geç oluyor kimi daha erken, çok sınıf olması ileri çok karıştırıyor.

Int: Doğru. Toplamda kaç öğrenci var hocam.

T1: Dönem başında kaç öğrencimiz vardı bizim? Bizim şu an 216 öğrencimiz var. Ama şu an 2500-2600 toplamda.

Int: Dönem başında?

T4: Dönem başında, 3500 falandı

T3: Mesela biz konferanslardan da bazı üniversitelerin test ofisleriyle tanıştık, “500 öğrencimiz var” diyorlar. Yani biz her seferinde çok büyük ölçekli sınavlar yapıyoruz aslında. Özellikle en yoğun eylülde.

T2: Bir muafiyet yapıyoruz. Eylüldeki en kapsamlı.

T1: Bunların hepsinin şeyini düşünün, optiği, zarfı, güvenliği, taşınması. Teknik sorun dedi ya Dilek, çocuklar bir listening sınavına girecekler teknik bir arıza çıktığında da biz ilgileniyoruz. Biz teknik personal değiliz ki baktığınızda. Niye CD’yi, USB’yi, bilgisayarla ben uğraşayım. Ben de bilmeyebilirim, ben bunu bilmekle yükümlü değilim ki. Ben de belki illiterate ım teknoloji kafasında.

T2: Sen, böyle bir eğitim verilmedikten sonra kendi kişisel çabanla belki biliyorsan ya da bilmek zorunda da değilsin aslında.

T3: Ve mesela bunların çoğaltılması hususunda kocaman bir makinemiz var, komple akıllı bir şey bunun teknik bilgisi şu an bizde mevcut ama biz bunu bilmek zorunda mıyız, ne kadar tartışılır.

T4: İşte o sekreteryaya düzeyinde işler hem enerjiyi öldürüyor hem yoruyor.

T1: Akademik anlamda da şey sıkıntı olabilir, bu sadece bizim okula has bir şey değil, öğrencilerin kültürel seviyeleri o kadar düşük ki biz bir parça ya da bir soru hazırlarken bunu bilirler mi, sürekli

genel olarak bir öğrenci profilindeki kültürel düşüklükten bir reading parçası seçmek bir cloze testteki konu seçimi çok zorlayıcı oluyor.

T3: Yaşlarına göre normal olan konular onlara soyut kalabiliyor. Öyle bir sıkıntımız olabiliyor.

T2: Soyut düşünme yetileri, evet bence öğrenci profili o, genel olarak. Diğer üniversitelerde de vardır büyük ihtimalle. Testing sürecini zorlaştırıyor.

T4: Ama işte MEB o kadar bozuldu ki, gelen öğrenci profili o kadar düştü ki hocam. Geçenlerde haberlere bile çıktı, 2-3 matematikle mühendisliğe gidiyor, böyle bir rezalet, mühendislik ya adı üstünde matematiğinin iyi olması lazım 2-3 matematikle artık mühendislik okuyanlar var.

T3: Genel bir sorun ama sorun nihayetinde, öğrencilerin ...

T4: Tabii teknik olarak baktığında fiziki koşullarda bizi çok zorluyor çünkü bu oda bize yetmiyor mesela, arkamızda gördüğünüz gibi bunları koyacak yerimiz yok. İkincisi güvenlik hususunda hocalar çok fazla girip çıkıyor buraya, o çok büyük sıkıntı. Çoğu zaman öğrenci de gelebiliyor, başka alakasız bir şey sormaya bir oda buluyor kapısını açıp, giriyor mesela. Ve bizim kapımızın üstünde test office yazıyor.

T1: Akreditasyon dolayısıyla koyuldu, birimlerin isimlerinin yazılması gerekiyormuş. Ama onun orada olması buranın güvenliği açısından dezavantaj.

T2: Bir de şey de var, yaptığınız iş motivasyon, motivasyon değil de konsantrasyon gerektiren bir iş. Bir kişinin gelmesi belki normal basit bir şey yapıyor olsan sorun değil, geldi baktı, yanlış yere gelmiş ya da atıyorum bir şey sordu gitti ama konsantrasyon gerektiren bir işte o ufaklık bir zaman belki senin yanlış yapmana o yanlışta fark etmesi geç.

T3: Özellikle belli anlarda, mesela örnek veriyorum ben orada shuffle yaparken birinin gelip bana soru sorması bütün konsantrasyonumu dağıtıyor. Ve biz burada genelde küme küme çalışıyoruz seviye arkadaşlarımızla birlikte. O anda birinin gelip bize soru sorması bütün dikkatimizi bütün motivasyonumuzu dağıtıyor açıkçası. Ama bir başkasına göre “ne olacak ki, geldim, sordum, gidiyorum” diyor çok zamanını almadım” diyor ama.

T1: Genelde devlet üniversitelerinde ofisler söylediğiniz sebeplerden çok tercih edilmiyor. Daha fazla kalıp çalışmak vesaire dediniz ya yazın geliyorsunuz teach and go yapamıyorsunuz. O sıra çok tercih edilmediği için bende yeni geldiğim için genelde yenileri böyle işte “sen de testinge geç” diye geçiyorlar. Ama şimdi o sistemimiz değişti müdür bey artık copy link açıyor, o ayrı konuda. Her isteyen değil, bir değerlendirmeden geçiyor. Mesela Sevim arkadaşımız öyle geldi.

T4: Kendi sekreteryası olan bir, ön hocalarında ilk soruları sorduğu ya da işte evrak işlerinin yapıldığı test ofisler olsak.

T1: Daire gibi ya da birim gibi tanınmıyor aslında mevzuatta, daire ya da birim olmayınca siz aslında görevlendirilemiyorsunuz yasal olarak görevlendirilemiyorsunuz. Bu da aslında uygulamada ihtiyaç duyulan ve yapılan bir şey. E o zaman yasal zemine niye kavuşturulmuyor.

T4: Karmadır, fotokopidir. Soru yazmakla uğraşsak.

T3: Sanırım İstanbul Üniversitesi'nde de öyle.

T2: Yapılmayacak değil, 1-2 kişi verseler bize.

T1: Yaşar'da aynen, onlar soruyu hazırlayıp teslim ediyorlar, geri kalan hiçbir şeyle uğraşmıyorlar Yaşar Üniversitesi'nde. Hiçbir şekilde biz muhattap olmuyor.

T2: Doğru Yaşar'da da öyle, normalde sekreteryaya düzeyinde çalışan kişiler var onlar ilgileniyor. Sizin sekreteryanız değil aslında onlar. Ama o işleri onlar yapıyor, sorumluluk verilmiş sadece. Boğaziçi'nde şöyle odası sizin odanız burada önce oraya gidiyor, sizin oraya geçemiyor.

T4: Şu karmayı öğrenci asistanı da yapabilir aslında.

T3: Gerekirse bu tarafa geçebiliyor, o çok önemli. Ben Boğaziçi'nde asistanlıkta yaptım, 5 yıl boyunca. Yüksek lisans dönemi boyunca. O zaman bile yani 20 yıl falan olacak o kadar zaman önceydi ki o zaman bile önce yabancı dilleri inanılmaz oturmuş bir durumdaydı.

T4: Onların da yine öğrenci sayısının fazla olması, beklentinin çok yüksek olması testlerde, en ufak hatada şey olması, biraz onları da zorlayan tabii yine sorunlar var.

T1: Motivasyon açısından İş yükünün eve taşması ekstra yorgunluk sebebi oluyor, bunlara vakit ayırmak yüzünden böyle oluyor.

T4: Mesela şeyle ilgili bende bir şey söylemek istiyorum, az önce bahsediyordunuz. Hocaların gelip gitmesi sanırım yanlış anlamadıysam, hocalar geldiğinde en ufak şeyde 'üff' yapıyor, o anda bizim motivasyonumuz şey yapabiliyor. Şey diyor "ne biçim sormuşsunuz" falan çok yaşadığımız vakalar değil ama arada bir böyle bir kaç kişi sitem ettiğinde bizde şey değişikliği oluyor, mesela şunları koymamız gibi değişiklik, "off bu ne ya bunu da mı biz yapacağız" bilmem ne gibi bazı şeylerle karşılaşınca o zaman böyle ister istemez bir saat konsantre olamıyorsunuz. Çünkü çok farklı şeyler canınız sıkılıyor, o tarz şeyler oluyor. Çünkü full içine odaklanmanız gereken bir durum ve en ufak bir şey çok çabuk dağıtıyor.

T2: Ve sınav sonrası süreçte de aynı şey, biz her şeyi muntazam bir şekilde hocaya verirken hocadan gelen şeyler o kadar muntazam olmuyor. O da bize ekstra bir yük mesela.

T3: Siz burada bir sistem belirliyorsunuz ama hocalar o sistemleri kendileri belirlemiyorlar ve yapmak istemiyorlar.

T1: Ama bunun toplantısı yapılıp anlatılıyor, böyle yapmanız gerekiyor ama yine de yapmadılar. Hoca peşinde koşturmak bence motivasyon kırıyor.

T4: Soru sormak daha kolaylarına geliyor, ilk elden bilgi edinmek, araştırmak yerine.

T3: Bazen gelen sorular da aslında onlara söylenen ya da toplantı da söylenen ya da handbookta olan bir şey bile aslında gelip soru mevzusu olabiliyor.

T2: Yani biz daha fazla çalışıyoruz bence daha fazla para almak zorunda gibi hissediyorum ben açıkçası. Çünkü ben diğer hoca gibi 25 saat işimi bitirip gitmiyorum ben ekstra mesaiye burada kalıyorum ve bunun yetişmemesi benim suçum değil. Hiçbir şekilde kendi suçum olduğunu düşünmüyorum.

T1: Bir de bu hoca 25 saat derse girdiğinde ek ders ücreti almış oluyor.

T2: Bizde alıyoruz ama biz 25 saatten fazla kalıyoruz. Yani 25 saat bitti gitti olmuyor bizde aslında.

T3: Evdeki çalışmalarınız hariç birde.

T1: Bir de onlar var.

T4: Mesela hocalar 13:30 da dersi bittiğinde gidebiliyor belki kağıt okuması vardır, essay okuması vardır, sınav okuması vardır. tabii ki o ayrı ama onun dışında biz, bu dönem ben hiç 13:30 da çıktığımı hatırlamıyorum en erken 17:00 de çıkabiliyorum. Artı evde çalışmalarımız var.

Int: Maalesef doğru.

T2: Ayrı ödenek olur herhalde, ayrı bir oda veya başkanlık olsak.

T3: hocam, uygulamada. İhtiyaç duyulan bir şeyin zaten uygulamada da, uygulama nedir, bir karar niye alınır altta bir ihtiyaç vardır. **(01:13)** boşuna yapılmıyor, bir ihtiyaç vardır yapılıyordur. Ama bu sefer şey kulağı şöyle tutayım derken biz şöyle tutuyoruz böyle de yürüyoruz. Böyle bir şey oluyor maalesef. Ek ödenek diyoruz ama bir sürü şeye bir sürü para harcıyoruz ama olması gerekenlere ya da mesela siz belki bir ikna yoluyla belki bir yöneticinin motivasyonu ile sizin kendi iç motivasyonunuzla belki böyle bir şey yapıyorsunuz ama çalışmıyorum deseniz, ek maddi bir şey yok manevi bir şey yok teşvik yapılıyor siz faydalanamıyorsunuz şimdiye kadar çıkan sorunlar. Atıyorum diğer hocalar dersi anlatıp gidebiliyor siz gidemiyorsunuz, yazın çalışıyorsunuz gibi şeyler. Çok saçma.

T4: En başından fiziksel koşul olarak baktığımızda da biz bir oda veya başkanlık olsak biz neden mesela fotokopiyi kendimiz çekelim veya onları niye kendimiz taşıyalım. Bütün başkanlık ve odaların zaten bunların bilgi belge işletme uzmanları da var, ulakları var. Baktığınızda o fiziksel külfeti de üstümüzde oluyor aslında.

T2: aslında çözülmesi gereken de aslında kolay çözülebilecek ama ya görmüyor ekonomik dediğimiz bazı sorunlara sığdırıyoruz bazı şeyleri. Sizin zaten bırakın fiziksel işleri sınavı hazırlama o kadar zor bir süreç ki bunun hazırlanmasından, relabılıtyinden, validitysinden.

T1: Mesela Nurgül şu an zarf yazıyor, 100 tane zarf yazacak şimdi. Zamanı değerlendirmem lazım. Onu şu an yapmak zorunda, düşününce bile şu an komik geliyor.

T3: Aslında şu çok teknik bir şey aslında materyal düzeyde bir şey aslında bu enerjiyi başka bir yere harcayabilirsiniz hocam ama.

T4: Onlar eve kalıyor genelde.

T1: Aynen evdeki iş gücümüze dönüşüyor onlar bizim.

3. Sınav süreçlerine yansıyan sistemik problemler nelerdir?

T1: Seviyeler arası Sömestr based olmamız.

T2: seviyeler arası geçişte problem var.

T3: Evet var. A1 A2'ler aslında gerçekten B1 olmuyorlar ikinci dönem.

T4: A1 B1'e geçiyor A2'lerde B1'e geçiyor. Öyle oluyor işte bilmiyoruz bizde. Kitapları yetiştiriyorlar programları bir şekilde, aynı kitabı okumuş oluyorlar ama tabii sonuçta öğrencinin seviyesi, başlangıçları bir değil.

Int: Ama yanlış anlamadıysam o zaman sene başında A1 oluyor öğrenci, bir grupta A2 oluyor. sonra eğitimleri alıyorlar.

T4: Sonra bahar döneminde birebir oluyorlar.

T1: Hocalardan aldığımız feedbackte bu yönde, öğrencilerin bazıları gerçekten ciddi anlamda bocalıyorlarmış bu dönem ama ilk defa uyguladığımız, bu yıl uyguladığımız bir sistem ve yani problemlili gibi duruyor B1 öğrencilerinde. B1 den B2 ye geçenlerde hiçbir problem yok gerçekten düzgün işledi o sistem haliyle ama işte A1 den B1 e atlatınca bazı öğrenciler.

T2: Ama orada onları ne kadar shuffle yapsan bile B1 kısmında ikinci dönem illaki bir şey olacaktır. Bazıları ilerde bazıları geride, ilerde olanda sıkılacaktır bu sefer, eksik kaldıysa.

T3: İlerde olanlara yönelik bir program oldu zaten, ağır bir program oldu. bu sefer arkada kalanlar oldu, curriculum çok yüküttü hocalar şikayet ediyorlar yetiştirememekten, yetiştiremedikleri için haliyle geride kalanlarla da ilgilenemiyorlar gibi genel hocalardan aldığımız bir feedback var yani.

T4: O zamanda sınavı hangi seviyeye göre hazırlayacağımız sorusu. A1 ile A2 yi ilk dönemin sonunda gördükleri gramer konuları açısından bakalım aynı yerlere geliyorlar. Hatta A1 ler A2 lerden bazı gramer konularını daha önden gördükleri bile oluyordu. Kitap farklılıklarından dolayı. Aslında Teoride aynı yere geldikleri varsayılıyordu.

T1: Ama işte gramer olarak sadece.

T2: Uygulamada olmuyor. Bir de şeyi soracaktım, ikinci dönem yanlış anlamadıysam sizde A1 ya da A2 hiç kalmıyor.

T3: Kalmıyor, B1, B2.

Int: En düşük B1 ve B2 şeyleri oluyor. İkincisi ara dönemde bölüme giden oluyor mu?

T1: İlk dönem B1deki öğrencilerimiz yeterlilik sınavını geçince.

T2: Evet, ara dönemde ocak dönemindeki İYS, proficiency girip geçebiliyor.

T3: Ama sadece B1 ler için.

T4: Bir de YÖKDİL'e girerlerse ya da TOEFL ya da uluslararası bir sınava getirirlerse geçebiliyorlar.

T1: Ama B1lerin de sadece belirli koşullarını sağlayabilenleri, herkes giremiyor. Belli notları tutturabilenler, 60 üstü.

T3: Bir devamlılık şeyi var yani.

T1: Çok fazla sınav hazırlıyoruz. Hazırlık programında proficiency yapıyoruz, midtermler, finaller, quizler. Bunların dışında UTE var mesela, mühendislik tamamlama var, yatay geçişler var.

T2: UTE var mesela, mühendislik tamamlama var, yatay geçişler var. Onlarda da yine proficiency gibi.

T3: Yabancı uyruklu öğrencilere bir şeyler oluyordu galiba, böyle ekstra sınavlarımız oluyor. Makeuplar var sınava giremeyen öğrencileri bir daha makeup yapılıyor.

T1: kalite ile ilgili bir süreçten geçiyoruz.

Int: Nereden hocam

T2: Pearson'dan

T1: Belgeleme, aşırı bir belgeleme oluyor.

T2: Bütün sınav arşivini istiyorlar, specificationları istiyorlar, sürekli her yaptığımız şeyi kağıt olarak raporlama. Mesela şu an bizden şeyi bile istediler, bu süreçleri fotoğraflamamızı. O dereceye kadar düşünün her şeyin belgesini istiyorlar, aklınıza gelebilecek her şey. O da ekstra bir yük o da ayrı bir sekreteryaya işi bizim için.

T4: peyderpey gelip söylüyorlar. Bugün şunu yapın, şu gün şunu yapın. Toplantıları çok fazla.

T3: O da kötü ama işte çünkü önceden bilsen sen ona göre planlarsın onu.

T1: okula göre her geldiklerinde toplantıda analiz yapıp şu eksik şunu yapın diyorlar galiba. Onlarında toplu bir listesi yok sanırım.

Int: Vardır da ya da vermek istemiyorlardır ya da bunu bir sürece yaymak istiyorlardır

T1: Olabilir.

T4: Bence çok bununla meşgul etmek istemiyorlar ama mesela biz oturup onun biliyoruz farklı farklı şeyleri var onları bizden istemediler. Daha sınav bazında bizimkiler, daha sınavla alakalı olan belgeler.

T1: Mesela mayıs ayının sonuna kadar müdür bey bizden bir belge istedi ve biz aşırı doluyuz şu anda.

T4: Spesificationları, iki kuru da. Onları istedi, şimdi onları ben yapmak zorundayım ben bir yanda. Şimdi ben bunu mu yapayım onu mu yapayım oldu. Yine bir sekreteryaya işi baktığınızda aslında.

T2: Kalite sürecinin bizim açımızdan olumlu olan bir yönü şey oldu aslında, biraz daha standartlaştı. o bir olumlu yönü.

T3: Belge yükü karşısında belki tam sirayet etmemiştir.

T1: Bir de sistem değiştiği için bizde aslında bizde oturtmaya çalışıyoruz. Belki seneye bu sıkıntıları bir tık daha az yaşayabiliriz bizde çok bilemiyoruz ama henüz oturtma aşamasında olduğu için daha da fazla hissediyoruz bu süreci.

Int: Dediğiniz gibi hem sistem değişikliği hem de akreditasyon süreci aynı anda denk gelince.

T1: güvenlik önlemleri yetersiz.

Int: Ne tür güvenlik önleminiz var?

T4: Kapıyı kilitleyip çıkmak.

T1: Güvenlik önlemi olsa şuradaki eski dolaptır.

T2: Evet, bizim burada bir tane çelik dolabımız var ama kapağı bozuk ama çelik dolap.

Int: Var mı hala

T3: Kaldırttık. Arşivimiz var. arşivin anahtarı bizde yok değil mi.

T4: Hayır bizde değil.

T3: Biz bile tutamıyoruz arşivin anahtarını bir tek fakülte sekreterinde güler hanımda yüksekokul sekreterinde var.

T2: Onu da biz sınavları göndereceğimiz zaman bizde eşlik ediyoruz, yerleştiriyoruz sonra anahtarı geri teslim ediyoruz. O anlamda bence güvenli.

Int: Peki buranın güvenliği açısından

T1: Buranın güvenliği kilitleyip çıkıyoruz

Int: Şu anda kilitli mi

T4: Değil, açık

T2: Diğer ofis kapıları ahşapken bu metal, materyalı tam bilmemekle beraber.

Int: Ciddi misiniz hocam

T2: Gerçekten evet, arada fark var

T1: Böyle **kanıtlanması** diye öyle düşündüler herhalde ama bunu da açmak çok zor olmasa gerek bir şeyler girdiklerini düşünürsek.

T2: Ama müdür bey bununla ilgili sistemi düşünüyordu. Şifreli giriş, yüz okuma, parmak, retinalı giriş. Güvenli çıkış. Akreditasyon sürecinde maddi olarak vardı düzenlik. o da onu olumlu düşünmüştü.

T1: Ama ofisin isminin yazıyor olması güvenliğe bir tehlike aslında yani eskiden o yazmıyordu mesela. O da bir güvenlik önlemiydi bence. Ama artık onun yazıyor olması bir tehlike.

Int: Peki yazmasının gerekliliği

T1: Akreditasyon süreci

T2: Gerekçesi ile ilgili bize bir şey söylenmedi

T3: Bazende hocalar, mesela öğrenci bizi sorduğunda buraya yönlendirebiliyorlar, o da bir güvenlik açığı. “Asena hoca nerde?” diyor “test odasında” diyor mesela hoca, buraya geliyor öğrenci. Öyle oldu bir kaç kere bana oldu.

T4: Derse girdiğiniz öğrencilerde gelebilir

T1: Derse girdiğimiz öğrencilere test ofiste çalıştığımızı söylemiyoruz

T2: Böyle küçük güvenlik açıkları

T3: Fotokopiyi falan güvenlik gerekçesiyle kendimiz hallediyoruz

T4: Ama mesela fotokopi çekmeye gidiyoruz. Hemen koridorun sonunda. Oradan oraya taşıyoruz biz.

T1: Ama o sırada size güvenlik eşlik ediyor.

T3: Bizi bir de öğrenciler bizi görüyor. Deminden beri sınav çekiliyor aslında

T4: Şu anda mesela sınav çekip çekip getiriyorlar, güvenlik eşlik etmiyor. Biz kendimiz. Ama öğrenci bizi görüyor elimizde o şeylerle, anlıyordur herhalde.

T1: Bir kere bir şey olmuştu, “aa quizi çekmişler galiba” falan diye söylemişlerdi. “Aa yok bu sizinle alakalı bir şey falan değil” demiştim öyle geçiştirmek için. Çok muhtemel çünkü buraya gidip gelirken benimde başıma böyle bir şey geldi yani.

T2: Mesela siz fotokopiyi çekerken bir öğrenci gelse oradan farkında olmadan.

T3: O da olur tabii, nasıl buraya giriyorlarsa oraya da giriyorlar yani.

T1: Baskı odası diye yazıyor nihayetinde üzerinde. Bu katta idari kısım olduğu için çok öğrenci burayla muhattap olmuyor ama idareyle işi olan öğrenci tabii ki buraya geliyor.

T4: Yolunu bulamayan öğrenci çat diye girebiliyor içeriye.

T3: Bizim şurada kapımız var mesela sürekli onu tıklatıyorlar.

T1: risk yönetimiyle ilgili Mesela şeyi söyleyebilirim, birkaç yıl önce soruların çalınması gibi bir durum olmuş ya da çalındığından mı şüphelenilmiş bilmiyorum artık tam detayını. Bu proficiency sınavı için o yüzden bir tane yedek bir sınav halihazırda tutacağız. Bir ve birden fazla olması gerekiyormuş aslında ama şu halde 1 hazırlayacağız. Bu dönemki görevlerimizin arasında o da var.

T2: Normalde bir sınav hazırlıyorsun bir de ekstra yedek olarak.

T4: Tabii yedek tamamen hiçbir şekilde dokunmayacağımız kenarda duracak, acil durumlarda şey yapabileceğimiz bir sınav olacak. Çünkü mesela en basitinden şey bile olabiliyor, atıyorum hatırlıyorum bir gün 2-3 sene önce tam sınav yapacağımız gün metrobüs kaza yapmıştı mesela metrobüsün kaza yapması bizim bütün öğrencileri etkileyen bir şey. O gün sınavı iptal etmek durumunda kalıyorsun. Mesela bazen şey oluyor hocalara vermiş oluyorsun, böyle 5-6 hoca almış oluyor paketi. E haber geliyor kaza oldu, o öyle olmuştu. E hocam geri getirin, belki hoca baktı ona. Bu tarz şeyler sınav güvenliğini bozan şeyler.

T3: Güvenlikle ilgili şeyde olabilir, bizim burada fotokopi makinemiz çok bozuluyor. Biz yukarıda başka bir binanın fotokopisine gidiyoruz. Oradan da müdür bey bize makam aracını veriyor ya da bir araç sağlıyor, biz yükleyip yine onları getiriyoruz mesela.

T2: Bir de eski sınavların arşivlenmesi hususunda yine ayrı bir harddiske daha atma durumu var, henüz yapamadık onu da. Böyle de bir planımız var.

T4: Öyle bir şey demişti müdür bey.

Int: Peki mesela yangın olsa burada

T4: İşte kapımız çelik.

T3: Yangın söndürücümüz var.

T2: Yangın olsa biz aslında şey düşünmüştük, ben onu müdür beye de ilettim. Cloud sistemine geçelim sınavlar cloudda olsun ama o zamanda cloudun da hacklenmesi durumu da var. O sıcak bakmıştı aslında en son ona. Hakikaten dediğiniz gibi her şey yine bir materyalde, komple okul yansa elimizde hiçbir şey yok. O yüzden de bir cloud sistemi sıcak baktı ama ondan sonra herhangi bir aksiyon alınmadı onunla ilgili. Bir cloud da olsa güzel olacak aslında.

T1: Zaten burası yansa her yer kağıt olduğu için herhalde çok hızlı yanar

T2: Çok güzel tutuřturur okulu

T3: Sınavdan bir gn nce allah korusun tabii de yangın ıksa burda sınav gn

T1: Bittik o zaman yani, okulu biz kapatıp gideriz. řimdi řyle bir řey var, bilgisayarlarda tutuyoruz ama diyorlar ki mail bile atmayın yine gvenlik aısından ama bu seferde tamam mail atma tut ki tek kopya benim bilgisayarımda, tut ki bilgisayar kt ne yapacaėım ben.

T4: Aynen yle, bir taraftan korurken bir taraftanda bařka bir risk doėuruyoruz aslında. Mailimizi atıyoruz bazen sınavların son hallerini.

T2: Bilmiyorum SYM nasıl koruyorsa sınavlarını neyle koruyor, bir dijital sistemleri vardır herhalde, yle bir řey satın alınıyordur byk ihtimalle.

T3: Ama o sistem yazılıyor, kendi sistemlerini yazdıkları iin hacklenmesi biraz zor olabiliyor.

T1: Onlar biraz daha řey ya offline alıřıyorlar, btn bilgisayarlar falan

T4: O da doėru.

T1: Hazırlık okulları dershaneye dnmesi. Daha mekanik oluyor.

T3: Evet ėrenciler genel olarak bir test sistemine alıřık oldukları iin ėrenmek deėilde gemek.

T1: sınavın sonucunda standardizasyonun nasıl saėlanması

T4: niversiteler arasında evet ok eřitsizlik oluyor. Mesela řeye baktıėınız zaman hocam, biz gramer soruyoruz ama Boėazii'ne baktıėımızda onlar gramer sormuyorlar onlar tamamen skills based olarak soruyorlar. O yzden de dediėi gibi nasıl olacak, ok farklı zaten sizde incelemiřsinizdir

T3: Soru tipinden tutun řeye kadar. bazı niversiteler listeninge daha az nem veriyor, bazı niversiteler writinge daha ok nem veriyor

T2: Aynen, sample İYS'lerine baktıėınızda Yani proficiencylerine baktıėımızda niversitelerin herkesin ok farklı řeylere odaklanıyorlar. O yzden bende katılıyorum arkadaşlara bir Boėazii'nin istediėi tip ėrenci yani proficient ėrencisi ile Yıldız'ın istediėi proficient ėrencisi bir deėil

T4: ortak bir curriculumun olması

T1: ben katılmıyorum. niversiteler arasında bir farklılık illa ki olmalı yani btn, her řeyi standart haline getirmeye kalktıėınız zamanda eřitliliėi bozmuř oluyorsunuz.

T3: Ders yapmalarını mmkn kılmak bence nemli, ėrenciler sadece sınavı dřnecekleri iin nasıl mesela lisede kabul etmesek de lise sondaki ėrenciler dersi dinlemek yerine test zmeyi

tercih ediyorlar, kapanıyorlar, soyutlanıyorlar bir şekilde. Aynı sisteme döner, çocuğu konuşturamazsın, derse meyil edemezsin.

T3: Bu sefer dershaneye döneriz, ÖSS sınavına öğrenci yetiştiriyor gibi öğrenciler bizden şey ister beni teste hazırla, bunu isteyecek bu sefer. İşte skill üzerinde durmaz amaç İngilizce öğrenmekten çıkar, ÖSS gibi “bana sınavı geçirt”e dönecek bu sefer.

T2: Şimdi bile hemen hemen öyle aslında, o kadar çok sınav odaklı ki öğrenciler

Int: Teşekkür ederim hocam.

4. Sınavlar hazırlanırken ne tür problemler yaşamaktasınız?

T1: Sınavlar hazırlama işte hammaliye işi. Sürekli fıtığa davet, ağır şeyleri indir kaldır. Sürecimiz çok yoğun.

Int: Nasıl hocam

T1: test developer midtermi ya da quizi hazırladığında önce partneriyle feedback alıyor. Önce beraber oturuyoruz, üniteler nereye kadar çıkacak curriculum yapıyoruz. Sonra bölümleri aramızda paylaşıyoruz, grammar vocabulary diye. Sonra materyal seçimi, hazırlanmaya başlamadan önce bir karşılıklı fikir alıyoruz, bu olur mu bu yapılır mı. Hazırlıktan sonra kendi aramızda, sonra diğer arkadaşlara gidiyor. Sonra düzeltip proofları müdüriyete gidiyor. Müdür yardımcısına, müdüre, CDU’ya, curriculum’a, native’e gidiyor. Onlar düzeltip. Native’e gitme aşaması, bunlar eş zamanlı gidiyor, bizden çıkması gerekiyor.

T2: Bizde zaman kısıtlılığı sebebiyle hepsine eş zamanlı gönderiyoruz. Önce CDO sonra native diye düşünüyoruz ama

T3: Zaman problemi doğru. Peki mesela müdüriyetten gelen revizyonları yapıyoruz. Bir daha bakıyoruz, beraberce. Bir ton nota, baskıya gidiyor

T4: Proof daha uzun bir süreç ve şöyle oluyor; bu şekilde bölüştürüyoruz kendi aramızda ve herkes hazırladığını ortak günde sadece o kısımla ilgili herkes feedbackini veriyor

T2: O gün mesela listening günü oluyor. O gün listening yapılıyor ve bu yaklaşık 6-7 haftalık bir proof sürecinden geçiyor

T1: Ayrıca Konu seçimi, çocukların genel cognitive levellerindeki sorunlarından dolayı

T3: İtem writing kısmında yukarda o metne uygun soru öyküsü yazmaya çalışıyoruz ya her soruya uymuyor mesela. Objective uygun soru çıkamayabiliyor

T4: Listening bulmak çok zor olabiliyor

T4: Geçende tartıştık. En sıkıntılı kısmı orası oluyor

T1: listening sınavlarını native okuyor. Ama bu da mesela nativemız okuduğumuzda tek bir aksan duymuş oluyor, başka bir aksan alternatifi olmuyor

T2: Sınavın ne zaman yapıldığı, ne kadar konu işlendiği, soru çıkamayabiliyor bazen. Çünkü çok az konu işlenmiş olabiliyor

T4: Tek bir soru üzerinde, bir kelime üzerinde bile şey yapıyoruz. Ama geri kalan süreci yine aynı şekilde ilerliyor, bizden çıktıktan sonra CDO'ya, müdüriyete, native'e gidiyor. O şekilde gerçekleşiyor

T2: Bazen zaman yetişmiyor. Özellikle bu dönem sonlarına doğru sınavlar üst üste binmiş oluyor. Aslında programımız var ama bir şekilde üst üste binmiş olduğu için yetiştiremiyoruz. Yetiştiriyoruz ama özel hayatımızdan okuldan sonraki kalan vaktimizden feragat ederek yapıyoruz bunu

T4: Görüyorsunuz sizinle mesela şu an oturup bile bir şey yaparken şunu da yapmak zorunda kalıyoruz. Genellikle durum böyle oluyor. Bizim şey şansımız hiçbir zaman olmuyor, burada ofisiz ama hadi gelin bir hava alalım, şurada bir kahve içelim

T3: Çoğu zaman yemek yemeye fırsatımız bile olmuyor, burada atıştırma şeklinde oluyor

T1: Yemekhaneye gittiğimiz günler sayılı. Zaman kaybı. Yemekhaneye gitmek bizim için 1 saat demek çünkü

Int: Niye

T2: Çünkü uzak, bir gideceksiniz yemeği yiyip buraya tekrar döneceksiniz. Atıştırma şeklinde ya da burada

T4: Biraz vaktimiz varsa gidiyoruz ama çok sıkışık durumlarda biraz şey oluyor. Yılda toplam kaç defa gitmişizdir

T2: 10 kere gitmemişizdir

T3: Bir memur gibi 12:00'de çıkalım 13:00'de gelelim. Bir de hocalarında ders saatine de biz dikkat ediyoruz, Burası hiçbir zaman boş kalmamalı, öyle bir şey var

T4: Bizim öyle bir şansımız yok

T3: Biz çoğu kez yemek yemiyoruz

T2: Bizim hocalarımız ve memurlarımız öğlen havuza gidiyor mesela

T3: 1 saat, öğle arası çalışamam diye bir şey yok

T2: 1 saat buranın boş kalması demek büyük intihar, hoca gelecek burada kimseyi bulamayacak

T4: Biz mesela atıyorum o an tek kalmışız tuvalete gitmişiz, hoca gelmiş bizi bulamamış, "test ofis kapalı" falan oluyor

T1: Söylemesi ayıp, kaç mesai saati burada belli değil

T2: Öyle bir şey yok. Nonstop ders var, 09:00'dan 16:00'a kadar. Öğle arası diye bir şey yok buradaki hocalara

T3: Tamam da idari açıdan düşünün, dersi düşünmeyin

T2: Hocalar aralarda geliyor ya da mesela öğleden sonra dersi olan hoca sabah gelip ders arasında gelip bir şey soracak oluyor. Demek istediğimi bilmiyorum anladınız mı? 12:00-13:00 normalde üniversitelerde öğlen arasıdır. Ama bizim ders saatlerimiz çok uzun olduğu için 13:30'da bitiyor

T1: Biz işte ona dahil değiliz, biz idari personel değiliz ya. Öyle bir şeyimiz olmuyor

T4: Allah var bak, o kadar yoğunum akşamları oluyor, haftasonu oluyor ama öğle aramı kolay kolay vermiyorum.

Int: Yeri gelen yerde idari gibi çalışıp, yeri gelen yerde akademik gibi

T1: 13:30'da bitse ne olacak? Diğer seviyenin dersi 12:00'de başlıyor. Arada hiç boşluk yok, o da bize yansıyor. Tabii bu sadece bir yönü, başka zaman açısından problem yaşıyoruz. Bazen öğrencilerin ilgi alanları demeyeyim de öğrencilere bazen soyut kalıyor hazırladığımız şeyler. Bu konuda da bazen problem yaşıyoruz, aslında normalde derste işlediği bir konu, üzerine konuşulan bir konu ama soyut kaçabileceğinden dolayı biraz daha basitleştirmeye çalışıyoruz

T2: Ya da şey oluyor, normalde diyelim bizim bir gün boş günümüz olabilir. Ama mesela gelmek zorunda kalabiliyorum, iş yetiemediği için. Burada herkes özveriyle çalışmak zorunda, kendinden bir fedakarlık vererek ama o fedakarlığın karşılığı olarak ekstra bir şey almıyoruz. Para istiyorum. Yok para için demiyorum sadece

T3: Yok, belki yurtdışında mesela bir eğitim olabilir 1 haftalık

T4: Evet, öyle bir şey olmuyor işte. Müdür bey kendince şey yapmaya çalışıyor

T1: Favorlar göstermeye çalışıyor

T3: Türk Hava Yolları'nın eğitimi vardı, bu pilotlara sertifika alıyorlar ya ona gidecektik. İşte önceliği bize verdi, biz Dilek ile ikimiz gittik, ilk bizdik ondan sonra geri kalan kısım diğer hocalardı. Bu durum hocaları çok rahatsız ediyor, o ayrı bir konu. Ama o kendince böyle bir şey yapıyor

T2: Ama şimdi bu müdür gitse başka bir müdür gelse, bunun bir sistem içerisinde olması gerekiyor

T1: Hocalarda biz çok şey yapmıyoruz bir tek ofisten çalışıyoruz kafasına giriyorlar, müdür beyin bakış açısı biraz daha farklı, ofislerle ben bu işi yürüteceğim diyor. Sizde o zaman ofise gelin diyor, insanlar pek gönüllü olmuyorlar

T3: Şeyden de olabilir o biraz; ben buraya girerken müdür bey “sen burada çalışacaksın” dedi bana, “ben biraz düşüneyim” dediğimde, “yok çalışacaksın” olduğu için, belki onunda favorımı da yapıyor olabilir birazda. Ben pişman mıyım, değilim burada çalışmaktan. Ama öyle bir görevlendirmeye gelmek durumundan olabilir

T4: Bazen bir iyilik yaparsın karşılık beklemek gibi bir şey değil, tamam iç motivasyonun var yapıyorsun, intrinsik bir şekilde yapıyorsun ama insan bazen de bu kadar yapıyor ediyor ve sürekli risk altındasın aslında sürekli alert, tetikte bekliyorsun. Sınavda bu sorun olacak mı, olmayacak mı, güvenlik önlemi, bir problem olacak mı, metrobüs kazası oluyor seni etkiliyor, o oluyor seni etkiliyor. Bu kadar şeyi düşün düşün, ek çalış, ekstra işler yap ama yine

T1: Ama hocaların gözünden bakarsanız onların sekreteryası işi çok fazla bizim hocalarımızın, sadece bizim değil. Onlarda kağıt kürek işiyle çok fazla uğraşıyorlar; yok yoklamasını gir, yok o notunu gir. Zaten haftada 25 saat çalışıyorlar, boş günleri yok ve bir ton evrak işiyle de uğraşıyorlar. yani okulda genel olarak bir sekreteryası sorunu var zaten sadece bize nazaran bir şey değil o bence

T2: Sınav hazırlamada bir çerçeve kullanmıyoruz

T1: Curriculum olarak CEFR kullanıyoruz sanırım, bizde aslında curriculum ile beraber yaptığımız için tam isim konulmuş olmasa da bildiğim kadarıyla CEFR'a bağlı

T4: Ama tam öyle konulmuş bir isim yok

T1: Konulmuş özellikle etiketi yok. Şeyi söyleyebiliriz; kelime listemiz yok, bizim bence en büyük sorununuz o. Bir kelime listesi olmadığı için ortak bir liste seçip onları işte o var mesela bence. Sorun dediniz ya, vocab de sıkıntı var bence. Arkadaşlar katılır mı bilmiyorum da

T2: çerçeve konusu beni biraz aşıyor şu anda. Çok bir bilgim yok benim şahsen

T4: Bizim de yok

Int: Hocam sizin eklemek istediğiniz bir şey var mı

T3: Bir taraftan kulağım sizde ama yok

T1: öğretilenlerle sınavda sorulanlar arasında örtüşme sorun olabiliyor.

T2: Ama birincisi biz zaten kendimizde derse girip curriculumu takip ediyoruz. Hepimizin bir skill ve grammar dersi var. Birbirimizle koordinasyon içindeyiz. Biz curriculumla da beraber çalışıyoruz, o yüzden ben ondan hiçbir sıkıntı duymuyorum. curriculum ile ortak yapıyoruz beraber, hangi üniteler, nereden nereye kadar, şurayı alın, burayı almayın diye. Yani beraber çalışıp aslında sınav içeriğini belirleyip onun üstüne soruyoruz. Çocuk görmediği hiçbir şeyden muaf tutulmuyor

T3: Target kelimelerden soruluyor sadece, kelime bilgisi. Vocabulary de seçenek hazırlanırken bile öğrendiği kelimeler seçenek olarak kullanılıyor, o derece

K2:Evet zaten derste öğretildiği gibi sormaya gayret ediyoruz

T1: Şöyle bir şey oldu aslında hocam; işin aslına bakarsanız dedim ya buradan önceki bizim şeylerimiz hep mühendislikti, yani bizim ilk defa ELT'den bir müdürümüz oldu. Ve biz bu standart edinmeyi, kalite edinmeye çok yeni başladık. Bir şeylerin isimlendirilmesi, adlandırılması süreçleri daha çok yeni başladı. O yüzden evet, tam olarak onlarla ilgili kapsamlı bir konuşma çok daha yeni başladı

T4: specification kullanmada ya da oluşturmada bir problem yaşıyoruz

T1: Quizlerin başlamasıyla alakalı olabiliyor, mesela bu dönem bize ilk quizimiz geçen dönem grammardı ve A1 seviyesinde grammar bir şey görmemişlerdi ki. Onu test yapabilmek çok zor olmuştu mesela. Belki onların biraz daha üzerinden geçilmeli. Quizlerin sırası, hangisi skilli ne zaman ölçeceği ya da dönemin başında bir listening quizi yapmak çok anlamlı olmayabiliyor. Bir şey görmemiş oluyor öğrenci çünkü. Şu sınavımızda böyle bir sorun yaşadık mesela; bu bahar döneminde quizde while listening sorduk ama çocuklar note taking eğitimi alıyorlardı, o note takingden bir şey sormadık çünkü dönemin çok başıydı, note takingde çok bir şey görmemişlerdi o yüzden geçen dönemin objektiveleri üzerinden bir quiz hazırladık. Bunların hepsi dedik ya, sınav sistemi değişti, eğitim değişti.

T4: Oturmadığı için henüz, gelecek sene herhalde bu problemi yaşamayız diye umut ediyoruz

T1: benchmarking yapmak istedik ama olmadı. bir üniversiteyle Boğaziçi'yle atıyorum, protokol yapmak ya da İstanbul Teknik Üniversitesi, onların hazırlığıyla bizden 10 öğrenci seçiyorsunuz onlardan 10 öğrenci, bizim sınavı onlara onların sınavı da bize. Müdür bey bununla ilgili bir şey düşünüyordu ama olmadı. Ben şey düşünüyordum siz gelmeden. Testingle ilgili bir sürü data geçiyor elimizden, yayın mı yapsak, ben doktora başvuracaktım. Sonra vazgeçtim çünkü bir hocam önce

“yayın yap boşver doktorayı” dedi. bizde Sibelle konuşuyoruz, o konu çok mantıklıymış, hatta Boğaziçi’nde yine Aylin hocayla da yapabiliriz

T2: Bizde yok sanırım, öyle bir uygulamamız yok

T3: İş akış şemasının net olmaması

T2: Bir genel takvimimiz var, baskı, şu, bu.

T3: Sen “A1 mi istiyorsun, sen A2 mi istiyorsun” ona göre kim kiminle ne yapmak istiyor ama İYS için yani proficiency için değişiyor. Şu yeşille yazılanlar da mesela ben bu İYS’de bunu yapayım sen bir sonrakini al. Kendi aramızda böyle yapıyoruz, öyle strict bir şey yok. Bu sorun olabiliyor bazen.

T4: Ama bütün prooflarda herkes birbirinin sınavını görüp mutlaka herkes birbirinin sınavını çözüp, tek tek geri dönüyor. Yani sınavımızı kendimiz çıkarıyoruz o seviyede bölüştüğümüz arkadaşımızla ama diğer bütün arkadaşlarla beraber yapıyoruz

Int: Şöyle; ikiniz bir seviyeyi paylaştınız sadece ikiniz görüyorsunuz ama değil mi hocam

T4: Yok hayır, herkes görüyor. Biz önce kendimiz diyoruz işte şu bölümleri yapacağız, bölümleri falan yapıyoruz, sonra bitirdikten sonra kendi düşüncelerimizle feedback veriyoruz bir orayı düzeltiyoruz. Sonrada arkadaşlarımıza dağıtıyoruz, onlarda düzeltiyorlar bir feedback veriyorlar.

T2: herkes aynı anda feedbackini veriyor, öyle düzeltiyoruz.

T4: daha fazla zaman kaybı yaşıyoruz.

Int: Diğer görevlerde paylaşım var mı

T4: Kim denk gelirse o anda

T3: mesela, ben şu an optik yaptım, Sezgin zarf yazdı bir dahakine ben zarf yazarım o optik yapabilir

T2: Yani bu benim işim değil, beni ilgilendirmez demiyoruz yani. Kim müsaitse.

T3: Bugün baskıyı kim yaptı ertesi gün paketlemeyi başka biri yapıyor

T1: Mesela dosya var, büyük dosya taşınacak bir yerden bir yere

T2: Çok büyükse görevli çağırıyoruz. O gelene kadar kendimiz yapıyoruz

T4: Biz yapıyorduk yani birçok şeyi

T3: Şu karşı oda komple arşivdi, sınavlar üst üste biz onların kendimiz paketledik tek tek o koca kolileri. Mesela bizim bölüm başkan yardımcımız arşivi düzenledi, kız bir hafta rapor almak zorunda kaldı. Bizde destek elemanlarını çalıştırmakta biraz zor oluyor çünkü öyle bir sıkıntımız var. Evet, fiziksel desteğe ihtiyacımız oluyor. O da şey değil mesela destek elemanının bu okulda bir ton işi var. Adamcağız 7/24 emre amade burada beklemiyor ki. Onu bekleyene kadar hadi kendim yapayım diyorsun mecburen

T2: Aslında şöyle; curriculum ve birimlere bir destek elemanı ayrılrsa, sekreteryaya ve destek elemanı

T4: Aynen öyle

T1: Soru havuzumuz yok.

T2: Zorluklardan benim aklıma ilk geleni zaman. Zaman yok, öyle bir zamanımız hiç olmadı oturup soru yazacak. Spesificationları, iki kuru da. Onları istedi, şimdi onları ben yapmak zorundayım ben

bir yanda. Şimdi ben bunu mu yapayım onu mu yapayım oldu. Yine bir sekreteryaya işi baktığınızda aslında

T3: Sadece çünkü soru hazırlamak değil onlar bizde çok fazla yere gidip prof alıp geliyor, onlardan dönüşleri de uzun bir zaman alıyor. Ekstra bu işler ve diğer işler var, o yüzden öyle bir şey olmadı.

Int: Normalde item bank yapılırsa daha herhalde iyi.

T4: Tabii daha iyi ama olamıyor şu an için. Bir de şeydi test office in sirkülasyonu da çok fazla olduğu için gelen insanların zaten çok, sürekli değişti içinde herhalde öyle bir şey. Bundan sonra düşünebiliriz

T2: Aklıma takılan soru tipleri değişti mesela, IYS'nin şeyi 2 aşamalı geçtik mesela, normalde note taking yapmıyorduk şimdi note taking yapıyoruz, soru havuzumuz olsa bile nasıl kullanacağız.

T1: Kaynak sürekli değişiyor.

T4: Kaynakta çok sık değiştiği için.

T3: Biz sistem değişikliğini gittik biraz da aslında onun da tabii etkisi oldu. Geçen senenin sınav türüyle şimdi farklı.

T4: Bir de her sene kitap değişiyor.

T1: Bir de şeye geçtik, o belki bir adımdır. CD kullanıyorduk normalde sınavlarda tek bir CD şeklindeydi bayağı yordu bu süreçte bizi ama nihayet USB ve bilgisayar kullanıma geçebildik. Bu bizim için büyük bir adım.

T2: Aya çıkmış gibi.

T3: Hocalarla ilgili sorun. Yani bizde burada cd çalar çalışmıyor diye, bize çağırıp fişe takmamış hocaları biliyoruz. Yani o yüzden bu sadece bizimle bağımlı bir şey değil teknoloji, dijitalleşme. Çocuklar okulun internet sitesinden girip sınav hakkında bilgi edinebiliyorlar, handbookları var ayrıca online olarak da var oradan da indirip bakabiliyorlar.

T4: Fotokopileri de biz saklıyoruz.

T1: hazırlama açısından geçerlilik-güvenilirlik sorun olabiliyor.

T2: Rater olayı yok bizde, content validity sağlamaya çalışıyoruz

T3: Çalışıyoruz, genelde homojen bir şekilde hazırlamaya çalışıyoruz konuları. Eksik kalan konularımız örnek veriyorum closest meaning sorularıyla halletmeye çalışıyoruz

T4: Öncelikle konularımızı çıkarıyoruz, hangi konular işlenmiş, hangi aralıkta neler işlenmiş şeklinde. Bunlara konu başlıklarını belirledikten sonra çıkardığımız soruları ona uygulamaya çalışıyoruz. Yani aslında entegre geçmiş oluyor. Onun dışında eksikiz

T1: Bir de şu şekilde de sağlamaya çalışıyoruz; mesela homework assignment verilmiş bir kelime seçerken oradan da bir kelime de alalım, writing bookletten geçende bir kelime alalım, listening kitabından da main course tan da alalım şeklinde her şeyden bir şeyler eklemeye çalışıyoruz. Ama ondan önceki scope belli zaten o scope un içinde bir şeyleri sağlamaya çalışıyoruz

T2: Şuradan bakarsanız, en son ya da şöyle göstereyim; bir corpora çıkarıyoruz bu bizde yok kendimiz hazırlıyoruz. Bu konuda ben şey yaptım. Başlıkları, konu başlıkları, bütün o dönem içerisinde gördükleri her şey, contemporary topics listening kitabı, reading kitabındaki

kelimelerimiz, assignmentlardaki kelimeler burada, writing packte geçen kelimeler, reading assignmentlarda geçen kelimeler bunları alıyoruz ve hazırlarken homojen bir şekilde olmasına gayret ediyoruz. Şuradan bakabilirsiniz, şu an gerçi bunu en son işaretlemedim ama. Genelde ünitelerin arasında da bir homojenlik sağlamaya çalışıyoruz bu şekilde. Bu aynı zamanda grammar konularında da var. Ama bu objektivelerimiz var burada, objektiveler belirleniyor. Objektiveler ve grammar pack işlendiği şeklinde sormaya çalışıyoruz. Nasıl time expression verirken bile packte veya grammar konusuna geçmiş haliyle, oradan verilmiş olan time expression lardan kullanmaya çalışıyoruz sorarken. Böyle bir dökümantasyonumuz var

T3: Gerçekten zaman kalmıyor dememizin sebebi böyle işler. Sadece oturup soru yazmadığımız için

T4: Bunca zaman çünkü burada dökümantasyon yapılmamış. Bir de öyle bir sıkıntı var. Önceki dönemlerden bize miras kalan öyle bir şey olmadığı için

T1: Daha sistemli bir çalışma gerekiyor. face validity bile bakamıyoruz.

T3: Face validity neydi

Int: kısaca şekil, format

T4: Format olarak bir dil sınavı formatında gözüküyor zaten. Bir de zaten soru tiplerinde ekstra materyallerde gördükleri, haftalık assignmentlarda gördükleri, sorunun kökü bile değişik kök yazmıyoruz yani derste geçen soru kökü bizim diğer sınavlarda geçiyor

T1: inter veya intra reliability açısından Yanlış bilmiyorsam öyle hiç birileriyle değerlendirmiyoruz ya da kendi içimizde okuyupta sen kaç verirsin onları şey yapmıyoruz

T2: Writing için daha çok yapılıyor o

Int: formative assessment yapılıyor mu

T3: Neydi tam olarak formative assessment. Jargondan uzak kaldım şu an

T1: Yeterince yapılmıyor. Ağırlıklandırmaya biz karar veremiyoruz.

T2: Hep beraber toplantıda konuştuk, müdüriyet, CDO, testing. Hep beraber toplantıda konuştuk. Sistemi yeni belirlediğimiz için toplantı yapmıştık, orada konuştuk. Ancak son onay aslında müdüriyette oluyor o zaman

T4: Evet, en son onlarda

5. Sınavların uygulanması sırasında ne tür problemler yaşamaktasınız?

T1: programa kabul şartları ile ilgili öğrenciler sınava girmiyorlar sınava girmeyince de random bir kura atılıyorlar. B1 seviyesinde A1 çocuğu oturuyor, A1 de B1 oturuyor. Ve o aralıklarda bence çok yeterli değil.

T2: Aralıklar hiç yeterli değil. Bir de ikincisi bilerek hazırlık “yatmaya” kendi tabirleriyle yatmaya gelen öğrenci, sınava girip random bir puan alıp sırf hazırlıkta kalayım diye gelen öğrenci, placement zaten yok placementımız da proficiencynin de bir anlamı olmuyor.

T3: Bir şey diyeceğim, öğrenci çokluğunu da yine şey yapabiliriz. Her sene kontenjan artıyor çünkü.

T4: Öğrenci sayısının fazla olması

T1: görevlendirme konusunda mesela bir şeyi muntazam verip, bizim bir yönergemiz var; biz sınavları o şekilde geri istiyoruz. Sıralanmış, imza listeleri tamamlanmış, imzalanmış şekilde. Hatta biz şeyde yapmıştık; İYS için, checklist hoca için, takibi kolay olsun diye. O checkliste yarım oluyor, ya kağıtlar sıralanmamış bir şekilde geliyor, çoğunun içinde signature disk çıkmıyor. Hocalarla o şekilde sıkıntı yaşıyoruz ve onun muhatabı biz miyiz, hocalarla

T2: Görev sınıfıyla da ilgili oluyor mesela ben burada girmek, bizim 2 tane bloğumuz var; bir blok oradaki sistem farklı, eski sistemden farklı, ses sistemi

T3: orada görev almak istemiyorlardı. Ben o sistemi bilmiyorum diye mesela

T4: Değiştirilmesini talep ediyor ama bu yüzden bizim tartıştığımız oldu mesela birkaç hocayla. “Ben bu sınıfta girmeyeceğim bununla değiştireceğim” diyor zaten orada 90 tane hoca geldiği için bizim “tamam hocam siz burada girin, şurada girin” demek o an çok kontrol edilebilecek bir şey değil

T1: Şu sistemde rahatladık, buraya artık yerleştiriyoruz burada hoca hiç bizimle muhatap olmadan kağıtlarını alıp gidiyor ve tekrar kağıtlarını buraya bırakıyor. Ama ondan öncesinde sınıf mevzusunda; bir de şeyimiz var ,yönergede sınıfta bulunma saatleri var sınavdan şu kadar dakika önce sınıfta bulunacaksınız ama hala son dakika kağıt almaya gelen oluyor. Şey mesela sıkıntı o durumda; ben burada o hocayı not edip bölüm başkanlığına bildirmem bekleniyor benim. O zaman hoca buradan şikayet edildiğini biliyor, bunu böyle bir proses olduğunu bilmiyor. Sanki biz “hocada almadı hemen bölüm başkanlığına...” ama bu bir hiyerarşik süreç. Ama o gelip buraya beni şikayet etmişsiniz moduna giriyor

T3: Muhatap siz kalmış oluyorsunuz, tartışmayı siz yaşamak zorunda kalıyorsunuz

T2: Evet, bir tane optik çıkmıştı hocanın kağıdının içinden, yok. Bizde bunu bölüm başkanlığına söyledik, “optiği bir öğrencinin yok” diye. Ama şikayet etmiş oluyorsunuz. Hoca buraya geldi işte “ben bu kadar senelik hocayım, öyle bir şey yapmış olamam” bir şey bir şey. Ama yapmış işte öteki kitapçığın arasına sokmuş onu. Olabilir, olsa possibility bıraksa hocalar olabilir dese bizde ona göre tavır alırız. O tarz şeyler teknik süreçlerde sıkıntı yaşıyoruz

T1: Bu arada sınavların düzgün uygulanmasıyla ilgili şöyle bir uygulamamız var; şu an o başlığı hatırlamıyorum var mı ya da geçecek mi daha. Checklistimiz az önce bahsettiğimiz, hocalarımız adım adım yapması gereken her şey var. Bunların Türkçesi de var, Türkçesini de veriyoruz asistanlar vs. geldiği zaman, çerçeve görevleri geldiği zaman. Bu da yönergemiz; her şey açık ve net bir şekilde adım adım before the exam, during the exam, the end of the exam şeklinde hepsi veriliyor

T3: Altta da numaramızı yazıyoruz, herhangi bir sorunda sınav esnasında bizi arıyorlar. Ama bu sınavla alakasız mesela tuvalete gidecekse de biz gidiyoruz, hoca ‘tuvalete gitmek istiyorum’ dediği zaman ben onun sınavına gözetmenlik yapıyorum. Kat görevlisi koyamıyoruz. Ama o da benim görevimin mesela

T4: Yedek bırakıyoruz, gelmeyen oluyor mesela yedekleri gönderiyoruz. Hoca çok olunca raporlu sayısı çok, izinli sayısı çok

T3: Doğru, doğru o da var

T2: Hasta vs. olma durumu var. Bu da sınav sonrasında bize hocalarımızın vermiş olduğu feedbackler. Buraya yazabiliyor. Ya da bir sıkıntı olduysa, yanlış optik veriyse çocuğu falan buraya not ediyor

6. Sınavların değerlendirilmesinde ne tür problemler yaşamaktasınız?

T1: Sınavların değerlendirilmesi çok fazla optik, optiği de biz okutuyoruz bu arada

Int: Hadi canım

T1: Bayağı alıp tek tek

T3: O baskı yaptığımız oda, odayı biz size gösteririz şimdi. Orada Kurtuluş Savaşı'nın 2 askeri, her gördüğümde de moralim bozuluyor, öyle bir resim yapıştırmıştım. Optiği de biz okutuyoruz. Optiği sonra bilgi işleme gönderiyoruz. Elektronik formatta çıkan o sonucu kaydedip, onlarda bize not olarak geri döndürüyorlar o sonuçları. Onlarda ayrı bir şey, artık biz göndermiyoruz ama bölüm başkanlığına verdik o işi çünkü o hiyerarşik bir süreç biz mi onun yazışmasıyla uğraşacağız diye

T4: Ama optik okumayı biz yapıyoruz

T2: Bizim midterm ve proficiencilerimiz optik

T3: Optik ve writing var

T4: writing süreci oluyor tabii. O da ayrı o da hocalarla bizim aramızda geçiyor

T1: Sınav sonrası da o çok sıkıntı, burada sürekli hoca trafiği. O geliyor kağıt alıyor, o benim birinci okumam neredeydi, ikinci okumam. Ben mesela Dilek ile partnerim, ben okuyacağım sonra Dilek okuyacak. Ben kağıdımı getirmiyorum Dilek beni bekliyor saatlerce, ben kağıdımı getireceğim diye

T2: Ve bize bazen kızabiliyorlar tabii o stresle, bir de deadline'ı yakalamaları zor oluyor hocaların bazen. Deadline'ı geçiriyorlar

T3: Deadline'ı çok aksatıyorlar, biz mesela burası sürekli bir sirkülasyon hali değil, bizim buradan sınavı toplayıp arşive indirmemiz lazım ki yeni sınavı yerleştirelim. Ama hala getirmeyen hoca oluyor sınavını, biz arşive gönderemiyoruz sınavları. Sürekli hocayı bekliyoruz getirsin diye

T2: İşte bir başkasına bağımlı çalışmak zorunda kalıyorsun

T4: O sıkıntımız oluyor, onunda takibini biz yaptığımız için hocayla da biz muhatap oluyoruz, "hocam getirmemişsiniz sınavınızı" genelde Asena arıyor hatta

T3: Ama genelde başka bir iş yaparken bir yanda da onu yapmak, en zor, sıkıntılı tarafı da o takip etmek. Takip etmek basit mi, bence basit bir olay. Ama şöyle basit değil, basit olmayan tarafı da; sen başka bir şey yapıyorsun o sırada onu yapmak zorunda oluyorsun. Unutsan ne olacak

T4: Öğrenci fazlalığının sebebiyle ama yani o süreçte hocaların okuma süresi falan hepsini hesaba katıp, bir de öznellik demiştik ya

T2: Uyguladığımız popülasyon kalabalık olduğu için optik oluyor. O yüzden reliabilityi aslında o şekilde sağlıyoruz. Writingde de first marking second marking var, hocalar bağımsız bir şekilde aynı öğrencinin kağıdını 2 kere değerlendiriyor. En sonunda bu 2 notlar yan yana getirildiğinde arada 4 ve üzeri fark olursa hocalar bir araya gelip bir anlaşılmaya varıyorlar, “bu çocuğun notu şu olsun ya da bu olsun” şeklinde. Eğer anlaşamazlarsa hocalar, onlar ayrılıyor third markinge gidiyor. CDO ve testing olarak da biz third marking konusunda destek veriyoruz

T1: Test hazırlama sürecinden sonraki standardizasyon süreci de sıkıntı olabiliyor. Yine çok fazla hoca ve çok fazla öğrenci olduğu için alternatif cevap çok çıkıyor, hocaların notlama sisteminde çok farklılıklar görülebiliyor.

T2: Özellikle writing gibi.

T1: Aslında writing için öncesinde bir toplantı yapılıyor, puan aralıklarına uygun kağıtlar seçiliyor öncesinde, örnek kağıtlar gösteriliyor “bu öğrencinin alabileceği puan aralığı şudur, şu öğrencinin budur şeklinde”. CDO ile, curriculum ile birlikte biz seçiyoruz, belli puan aralıklarına göre örnek alıp tekrar. Notlarına da beraber karar veriyoruz, ortaklaşa. Notlarını öncesinden tartışıyoruz, karar veriyoruz. Sonra toplantıda bunlar hocalara gösteriliyor. Hocalar toplantıda okuyorlar bu arada. Biz işte “bu kağıt bu puan” demiyoruz. Onlara okuma süreci veriliyor okuyorlar tartışıyoruz üzerine. Ama herkes genelde bitsede gitsek modunda olduğu için öyle çok bir tartışma ortamı olmuyor

T4: Doğru

T2: Diğer sınavlarda da zaten optik olduğu için daha doğrusu midtermler optik olduğu için ondan bir sıkıntı olmuyor. Quizlerde de biz şey yapıyoruz; açık uçlu olabilecek şekilde olan sorularda olabilecek cevapları mutlaka cevap anahtarlarında veriyoruz. Ve örnek veriyorum mesela bazı sorularda ‘no partial point’ şeklinde özellikle belirtiyoruz ki standardizasyonu sağlayalım diye

T3: Soru hazırlarken açık uçlulardan çok sıkıntı yaşıyoruz. Standardı sağlamada çok zor oluyor.

T4: Orada da avantaj dezavantaj çıkmış oluyor. İşte o yüzden bizde bundan kaçınıyoruz, bilinçli olarak.

Int: Neden kaçınıyorsunuz hocam?

T4: writing ve açık uçlu sorulardan. Bu kadar öğrenciyle.

T1: sınavların değerlendirilmesi sonrasında istatistiksel analiz yeterince yapılmıyor

T2: Bazı sınavlarda yapıldı özellikle yeterlilik sınavlarında yapıldı. Bilgiler paylaşıldı

Int: Peki bu konuda sonuç nasıldı hocam

T1: Öğrencilerden şikayet gelmişti o zaman. Bizim okulumuzda şöyle facia durumları olmuştu; öğrenci sınav çok zor deyip rektörlüğe şikayet edip, rektörlüğün buraya yeniden sınav hazırlattığı durumlar oldu. En son yine böyle bir şikayet oldu ve bu sefer müdür bey bu konuda çok dirençli çıktı, bütün analizlerini yaptırdı sınavın ve sınav moderate seviyede bir sınav çıktığı için rektörlüğe

direnebildi. O bu durumda bizim arkamızda durdu o konuda. Tamamen öğrencinin safsatası olduğu meydana çıktı bunun. O yüzden faydalı oluyor tabii, bunu yapmak gerekiyor

T2: Her sınav sonrası quizdir, midtermdir, sınıf ortalamaları paylaşılıyor öğrencilerle. Öğrenciler bu istatistiksel ortalamaları görebiliyorlar

T4: Bir de çocuklar, başarılı olan ilk 3 sınıfa sertifika veriyorlar. Quizlerde ya da midtermlerde. Başarıınızdan dolayı teşekkür amaçlı

T1: Proficiencylerde ise daha açık bir istatistik oluyor. Kaç beklemeli öğrenci girdi, bunun kaç geçti kaç kaldı, yeni öğrencilerin kaç girdi kaç kaldı, devamsızlıktan kalan öğrencilerden daha sonra sınava girenlerin kaç geçti kaç kaldı şeklinde detaylı analizlerde olmuştu ama sanırım bundan sonra daha yoğun olacak, bu ofisin kurulmasıyla birlikte daha yoğun bir istatistiksel bilgi olacak

7. Sınavların değerlendirilmesi sonrasında ne tür problemler yaşamaktasınız?

T1: Henüz bizde Self Access Center yok öyle bir sistem yok sanırım. Öğrenciler sınav formatını bilmediğinden şikayetler oluyor.

T2: Sadece Sınav örnekleri öğrencilerle paylaşılıyor, proficiency örneği atıyorum.

T3: Eski sınavları toplu olarak da satın alıyorlar öğrenciler.

Int: Nasıl yani

T3: Bilmiyorum bir şekilde fotokopi merkezlerinde var alıyorlar, nerden geldi oraya bilemiyorum. Eski sınav çıkmış soruları görüyorlar. 2010-2009 falan var ellerinde

Int: Allah Allah

T4: satılıyormuş herhalde bir dönem üniversite tarafından. Çünkü ODTÜ'nün de vardı ben hatırlıyorum galiba, onlarda bir dönem veriyorlardı.

Int: ODTÜ'de şey değil yalnız, satın almakla değil, kitap çıkarıyor.

T3: Bizimki de kitap şeklinde, okul satmıyor. Kitap şeklinde var okul satmıyor. bizde fotokopici kazanıyor.

T4: Orada bir de şey önemli tabii, ODTÜ'nün yaptığı ile burada ya da başka yerde yapılan iş ticari gibi kalıyor ama orada kurum öğrenciler kendilerini geliştirsın yapsın görsün mantık çok farklı.

T3: Bir de mesela sınav sonrası problemle ilgili şu olabilir; öğrenci her şekilde itiraz etme hakkına sahip. Kağıdına geçerli bir neden bile yazmıyor, sadece “aldığım nota inanmıyorum” gibi bir saçma sapan söylemler “kağıdımı tekrar değerlendirin” . Biz topladığımız sınavlardan çocuğun sınavını tekrar bulup, tekrar kontrol edip, bir sıkıntı var mı diye. Ve bundan kaç tane öğrenci şikayet ediyor

T1: gözle görülen problem itiraz eden öğrenci sayısının fazla olması ve itiraz ederken ki gerekçelerinin çok önemli olmaması. Bazen öğrenciler mesela gerekli bilgileri yazmayı unutuyorlar maalesef

Int: Nasıl yani hocam

T1: Öğrenci numarasını yazmıyorlar bir de listeden onu araştırıyoruz öğrenci numarası neymiş diye

T3: Bu bizim işimiz olduğu için direkt biz ilgileniyoruz. Tek tek açıp listeden bakıp bulmaya çalışıyoruz. Bunlar ufak tefek sorunlar aslında ama öğrenci sayısına bakınca

T1: Ama işte zaman alıyor sekreteryaya düzeyinde

T2: Kontrol etme söylemesi ayıp şey; siz yapıyorsunuz ve yaparkende, gerçi optik okuma olduğu için öyle bir hata olmaz ama. Optiğini de çıkarıp elde tekrar okuyoruz

T3: Optiğini kitapçığı ile karşılaştırıyoruz. Çocuğun optiği A mı kitapçığı A mı öyle bakıyoruz

T4: Hatta bir keresinde 600 tane itiraz olmuştu, 600 dilekçe

T3: Çünkü o zaman şey söyletisi çıkmıştı; sınav çok ağır, zor yapıldı

T4: Hocam o kağıtları tek tek bulacaksın oradan. Bulduk, çıkardık tek tek. Komisyon ayarlandı. Sonra çocuğun dilekçesine cevap yazıyoruz hata bulunamamıştır diye.

T1: Bir de onların hepsine tek tek cevap yazmak zorundasınız, o da ayrı yine zaman çok fazla zaman alıyor

T2: Bir de o da var. Dilekçe olduğu için bir cevap alması gerekiyor. “Maddi hata bulunamamıştır dilekçeniz şey olmuştur”.

T3: Maddi şeyi bile yazsan zaman alıyor sonuçta. Hepsine tek tek yaz, tekrar paketle, yerine kaldır. Zaman

T1: Ama o zaman çocuğun böyle bir lüksü var mı, iş bitti. Haklı gerekçesi olmadan sadece “tekrar bak” diyor bana, neye nazaran tekrar bakayım. “Aldığım not çok düşük tekrar bak” diyor. Öyle bir itiraz hakkı saçma geliyor bana açıkçası

Int: Peki bunu nasıl aşabilirsiniz

T1: Bu kadar itiraz hakkı vermesinler. Yine sekreteryaya işi bu, ya da yanımıza birini versinler çıkarmalarını o yapsın zarftan, ben kontrol edeyim o tekrar geri koysun

T2: Çünkü itiraz hakkını elinden almak çok demokratik bir olay olmayabilir. Ama tekrar kontrol edilmesini istiyorum derken bir gerekçesi olmalı. Öğrencinin keyfinden ziyade bir sebebinin olması gerekiyor yoksa öteki türlü 600 tane dilekçeyle uğraşmak doğru olmuyor

T4: Şöyle de bir sıkıntı var; mesela 60 ile geçebilecek 59 da kalmış. Onun derdi “bana 1 puan verin de writingim düşük geldi, ordan bana 1 puan verinde beni geçirin”. O itiraz sebebi bazen de o

T1: Öğrencinin öyle bir beklentisi olabiliyor

T2: değerlendirme aşamasında geçerlilik güvenilirlik ile ilgili hocalar belki speaking de ne kadar rubric olsa da hocaların inside ile hareket etmeleri çok fazla oluyor. Speaking de bence sıkıntı yaşanıyor. Daha öznel bir şey çünkü ve anında karar vermeniz gereken bir şey olduğu için

T3: Aslında rubric ler bu konuda yol gösterici ama yine de hocaların değerlendirmesi biraz daha anaç yaklaşılabiliyor bazı hocalarımız. Dolayısıyla o arada puan farkı olabiliyor. Bir de sanırım yanlış hatırlamıyorsam speaking de bir puan aralığı farkı yok. İkisinin ortalaması oluyor. Bu da açıkçası çok sağlıklı bir değerlendirme midir, o konudan çok emin değilim. Ama speaking değerlendirmesi hususunu CDO yapıyor biz onunla çok ilgilenmiyoruz

Int: Niye öyle hocam

T3: Bilmiyorum öyle

T1: Aslında şöyle; writing konularını beraber belirliyoruz ama speaking de biz bu konuda biraz daha pasifiz

T2: Aslında CDO dönem içindeki bütün materyalleri hazırlıyor ya hocam, öğrencilerin bütün speaking, listening materyalleri o hazırlıyor. Biz materyalleri o kadar erişimimiz ve zamanımız olmayabiliyor, sınav hazırlarken, speaking değerlendirmesi yapmıyoruz çünkü biz bu sınavlarda. CDO o işin içine daha dahil olduğu için çocukların speaking curriculum'a, onların yapması daha sağlıklı oluyor

T1: Daha sağlıklı bir işleyiş oluyor. Bir de sanırım birçok üniversitede CDO ve testing aslında entegre çalıştığı için böyle bir ayrım yok. Burada da aslında biz bir koordine şeklinde çalışıyoruz

T2: Bir çember gibi aslında

T3: Sadece adlarımız ayrı. Gerekli yerde birbirini besleyen

T4: Evet

T1: sınav sonuçlarını duyurduktan sonra orada ki sıkıntı öğrencinin genel argümanı hep zor olduğu yönünde, öyle bir sıkıntı yaşıyoruz ama onun dışında

T3: Genelde öğrencilerden bir direnç görüyoruz sınavlarla alakalı. Konu seçimleri olsun

T1: Genelde zaten konu açısından da çeşitlilik sunamıyoruz. Teknoloji ya da güncel konular yakalayamıyoruz genelde. Bu konulardan da eksi dönütler alıyoruz çocuklardan

Int: Dönütleri nasıl alıyorsunuz hocam

T4: Müdür bey şöyle bir şey istedi; hoca doğrudan gelip buraya size sözel bir yorumda bulunmasın. Bunu bir dilekçe şeklinde bir yazılı bir şekilde iletmesi daha sağlıklı diye düşündü. O yüzden hocalar doğrudan gelip artık bize bir şey demiyorlar, herhalde yazılı giden bir şeyde yok bize bir şey gelmiyor. Ama şöyle de bir durum var; biz burada hepimiz arkadaşız hocalarla. Gayriresmi bir şekilde zaten o dönütleri biz alıyoruz; nasıldı, neydi, öğrenci ne yorum yaptı. Bir de kendimizde derse girdiğimiz için çocuklara çaktırmadan soruyoruz sınavla ilgili geri dönüşleri ama official bir şey olarak

T3: Official olarak bir tek bu belge var, bu da sadece midtermlerde ve proficiencylerde var. Quizlerde öyle bir uygulamamız yok

T2: Ama bu belge sadece olumsuz yorum

T4: Tabii, normal

T1: Feedback hep olumsuz feedback varsa yazılıyor olumlu bir şey onlar yok

T2: Türkiye’de genelde biliyorsunuz olumlu eleştiri adına çok bir şey yok

T3: Genelde olumsuz feedbackte gelen şeyler; sınavın zorluğu oluyor

T1: Süreyle alakalı mesela 75 dakika verdiysek “bu süre biraz fazla bunu biraz azaltalım” şeklinde. O da sıkıntılı çünkü aslında hoca sadece girdiği sınıfa göre bir yorum yapıyor ama aslında yan sınıfın seviyesi aynı hazırbulunuşlukta olmuyor. O öğrenciye yeterken, o öğrenciye fazla geliyor. O standardizasyonu sağlayabilmek için onu yapıyoruz

T4: En alttaki öğrenciye düşünüp yapıyoruz

T3: Süre konusunda hep feedback geliyor zaten

T1: Sorularla alakalı zaten feedback gelmiyor genelde süreyle alakalı

T2: Sorularla ilgili de bazen oluyor ama bence o çok sağlıklı bir şey değil aslında çünkü hocanın sınav gözetmenliği esnasında bir soruyu değerlendirip bunun üzerine yorum yazması sağlıklı bir şey değil yani. Çünkü sınava bakmış demek gözetmenlik yapmamış demek oluyor aslında

T3: Hem öyle hem de şöyle bir şeyde var biz bir soruyu yazarken sadece o skilli değerlendirerek yazmıyoruz bunu grammarda o konuyu işlediyse de dahil etmiş oluyoruz. Ya da o soru tipini gördüyse dahil etmiş oluyoruz. Sadece hocanın o anda bakıp gördüğü sorunun altında çok daha fazla şey var. Onları düşünemedikleri için çoğu zaman bazen o konuda feedback alıyoruz

T2: Genellikle eski test ofisi çalışanlarından geliyor, en çok yorum onlardan geliyor. Normal hocadan pek yorum gelmiyor gibi bir şey hatta

T1: Yapıcı yorum oluyor onlarda. “Şunu şöyle yapsanız daha iyi olur” şeklinde yorumlar geliyor

Int: Pek, öğrencilerden

T1: Anketi yönetim uyguluyor sanırım

T2: Yönetim, aynen. Şu an online bir şekilde uygulanıyor eskiden optik formlarda. Şu an onu hatırlayamadım çok emin değilim ondan

Int: Peki, sizle paylaşıyor mu

T2: Aslında o anket hocalara yönelik, öğrenciler hocaları anketliyorlar, aslında bizi de anketliyorlar ama.

T4: Bozuldu birazcık bir sıkıntı çıktı, bizde çıkmadı da diğer hocalar zarflarını almış ya

T3: Şey oldu ya onun dışında da yapıldı; curriculum değerlendirildi “beklentileriniz nelerdir” vs. diye

T1: Bir de hocaların değerlendirdiği şey var, mesela curriculum yayınlıyor onu. Google anket gibi bir şeyden. Oradan hocalar curriculumu değerlendiriyor ama test ofisin değerlendirildiği bir şey yok.

T2: Ayrıca sınavların arşivlenmesi biz tarafından, hepsi biz tarafından yapılıyor.

T3: Yalnız hepsini birden konuşunca dünya kadar işini yapıyormuşuz. Parça parça ama

T4: Hocalar sınavları getiriyorlar. Evet, bayağı bir şey. Ben şimdi size şuradan göstereyim, Mehmet bey. Bakın şurada mesela en son yaptığımız quizler var, onlar paketlenip arşive

Int: Ama her şeyi siz mi yapıyorsunuz

T1: Her şeyi biz yapıyoruz

T4: Onları şöyle pakettiyoruz; bize çöp torbası veriyorlar. Çöp torbası veriyorlar, siyah bir tane ona koyuyoruz. Üstüne de hangi sınavın olduğunu yazıyoruz

Int: Ciddi misiniz

T2: Evet gerçekten. Sonra görevli geliyor, beraber arşive gidiyoruz, görevli söylene söylene onları yerleştiriyor. “Siz zayıfsınız siz daha iyi girersiniz” diyor, bir de onla bir tartış, sonra yerleştiriyor onları

T3: Çöp torbası, ceset gibi

Int: Tövbe estağfurullah. Ya karışsa o çöp poşet

T4: Yok, karışmıyor dönemi yazıyor senesi yazıyor

T3: Kurbanlık getirir gibi

T4: Türkçe yazıyor ki diğer çalışanlar anlasın diye

T3: Çöp poşetine koymamızın sebebi onu daha korunaklı bir hale getirmek. Ama imkanlar dahilinde

T1: Şey sıkıntı olabilir mesela ben onun sıkıntısını yaşıyorum; 2 gün sonra ben arşive insem o poşeti nasıl açacağım, yırtmam lazım ama onu

Int: 1 saniye, o arşivde o poşetin içinde mi kalıyor

T2: Evet burada tutamıyoruz sadece sınav. Ama etiketli. Midterm 1, B1 level. Ama bence sıkıntı. Orada poşetler oluyor, bir sürü poşet oluyor

T1: Ama arşiv hayalinizdeki gibi üst üste, üst üste değil. Açılan dolaplar var, olur ya devlet dairelerinde. Maalesef burada arşiv dediğinizde taşımak için çöp poşetinde taşıdın, ilgili çekmeyi açtın, içerisine yerleştirdin. Oranın revize edilmesi lazım.

T3: Çünkü dolaplar böyle açılıyor araları böyle ayrılıyor. Ajanlar filmlerde olur ya arasına girer öyle bir şey

T4: Ama onunla ilgili müdür bey aksiyon alacakmış çünkü akreditasyon sürecinde arşivle ilgili revizyon yapılması gerekiyormuş

T2: ne kadar sürede imha edilmesi gerekiyor belli değil

T1: belki bir motive edici bir teşvik gibi bir şey olabilir, belki atıyorum yasal bir zemin oluşturulup ne için çalıştığımızı.

T3: Belki kurumsallaştırılır.

T2: Kesinlikle. Belki dediğiniz gibi sekreteryaya görevi bir ayırt edilir.

T1: Bizim mesela Spor, Sağlık Kültür Daire Başkanlığı var. Bu olayı da sevmiyorum ama ya 2 senelik mezunlar çalışılıyor ya lise mezunları çalışıyor orada ama hepsinin emre amade şeyi var. Yok o belgeyi getir, bu belgeyi getir.

T2: Ama sen burada özel şeyin için istemiyorsun ki o kişiyi. Kullanacaksan da o kişiyi, özel bir zevkin için kullanmıyorsun canım, Allah Allah.

Yukarıda sorduğum hususlar dışında belirtmek istediğiniz başka bir husus var mıdır?

Int: Hocam ben size kartımı bırakacağım, eğer buna ekstra benim unuttuğum bir konu varsa mail atabilirsiniz ya da mesaj atabilirsiniz. Mesela buna soru olarakta “şu soruyu da eklersen iyi olur diyebilirsiniz” ya da “şu konu da gündeme gelmedi, unutuldu, söylemeyi unuttuk”. Konu olarak da soru olarak da, iki şekilde de ek olarak şey yapabilirsiniz.

T2: Őu an benim kafam biraz Őey, henüz odaklanamıyorum ama bakarız tekrardan size dönüt verebilirim.

T3: Anlaşılan herkes istifa edecek.

Int: Olur hocam, yoğunsunuz farkındayım. Çok teşekkür ederim valla çok güzel oldu.