REFERENCE TO TESTING PRINCIPLES AS AN INTERACTIONAL RESOURCE IN L2 TESTING AND EVALUATION CLASSROOM INTERACTION IN TEACHER EDUCATION CONTEXT

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MÜBERREM BERNA BAYDAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF ARTS
IN
THE DEPARTMENT OF ENGLISH LANGUAGE TEACHING

SEPTEMBER 2022

Approval of the thesis:

REFERENCE TO TESTING PRINCIPLES AS AN INTERACTIONAL RESOURCE IN L2 TESTING AND EVALUATION CLASSROOM INTERACTION IN TEACHER EDUCATION CONTEXT

submitted by **MÜBERREM BERNA BAYDAR** in partial fulfillment of the requirements for the degree of **Master of Arts in English Language Teaching, the Graduate School of Social Sciences of Middle East Technical University** by,

Prof. Dr. Sadettin KİRAZCI
Dean
Graduate School of Social Sciences _____

Prof. Dr. Nurten BİRLİK
Head of Department
Department of Foreign Language Education _____

Prof. Dr. Çiler HATİPOĞLU
Supervisor
Department of Foreign Language Education _____

Examining Committee Members:

Assist. Prof. Dr. Nilüfer CAN DAŞKIN (Chair)
Hacettepe University
Department of Foreign Language Education _____

Prof. Dr. Çiler HATİPOĞLU (Supervisor)
Middle East Technical University
Department of Foreign Language Education _____

Assoc. Prof. Dr. Perihan SAVAŞ
Middle East Technical University
Department of Foreign Language Education _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Müberrem Berna BAYDAR
Signature :

ABSTRACT

REFERENCE TO TESTING PRINCIPLES AS AN INTERACTIONAL RESOURCE IN L2 TESTING AND EVALUATION CLASSROOM INTERACTION IN TEACHER EDUCATION CONTEXT

BAYDAR, Müberrem Berna
M.A., The Department of English Language Teaching
Supervisor: Prof. Dr. Çiler HATİPOĞLU

September 2022, 183 pages

The growing demand for language teachers to be equipped with the necessary skills to conduct assessment-related activities as a part of their job resulted in the publication of a considerable amount of literature on language testing and assessment to understand the needs and expectations of education programs fully. However, there is still a gap in the literature regarding the investigation of how language assessment literacy of preservice teachers develops in and through interaction. Therefore, this study aims to fulfill this gap by focusing on the phenomenon of "reference to testing principles" in the context of language testing and evaluation course at an English Language Teaching (ELT) program by adopting a conversation analytic approach. The study draws on 12 hours of video-recorded classroom interaction data gathered from senior year ELT students and an ELT professor at a state university in Ankara, Turkey. Based on CA, this study has investigated the emergence of the phenomenon, reference to testing principles in different sequential environments to understand the functions it performs during the peer feedback interaction sessions in language testing and evaluation course. Furthermore, the study has uncovered the relationship between the

test types and testing principles oriented by the preservice teachers. The analysis has also shown that reference testing principles during the peer feedback interaction provides learning opportunities to develop language assessment literacy and skills necessary for language teachers. All things considered, the study reflects on peer feedback interaction in a higher education context and offers implications for L2 teacher education research and the development of language assessment literacy.

Keywords: Conversation Analysis, Classroom Interaction, Language Testing and Evaluation, Invoking Testing Principles, L2 Teacher Education

ÖZ

BİR ÖĞRETMEN EĞİTİMİ BAĞLAMINDA GERÇEKLEŞEN YABANCI DİLDE ÖLÇME VE DEĞERLENDİRME SINIF İÇİ ETKİLEŞİMİNDE ETKİLEŞİMSEL BİR KAYNAK OLARAK TEST İLKE VE PRENSİPLERİNE REFERANS GÖSTERME

BAYDAR, Müberrem Berna
Yüksek Lisans, İngiliz Dili Öğretimi Bölümü
Tez Yöneticisi: Prof. Dr. Çiler HATİPOĞLU

Eylül 2022, 183 sayfa

Dil öğretmenlerinin işlerinin bir parçası olarak ölçme ve değerlendirme faaliyetlerinin yürütülmesi adına gerekli becerilerle donatılmasına yönelik artan talep. Bu bağlamda eğitim programlarının ihtiyaç ve beklentilerini tam olarak anlamak için yabancı dilde ölçme ve değerlendirme konusunda önemli miktarda literatürün yayınlanmasıyla sonuçlandı. Ancak, öğretmen adaylarının yabancı dilde ölçme ve değerlendirme okuryazarlığının etkileşim içinde ve etkileşim yoluyla nasıl geliştiğinin araştırılmasına ilişkin literatürde hala yeterli derecede çalışma yoktur. Bu nedenle, bu çalışma, bir İngiliz Dili Öğretimi (ELT) programında yabancı dilde ölçme ve değerlendirme dersi bağlamında "ölçme ve değerlendirme ilkelerine referans gösterme" olgusuna odaklanarak bu boşluğu söylem çözümlemesi yaklaşımını benimseyerek doldurmayı amaçlamaktadır. Çalışma, Ankara, Türkiye'deki bir devlet üniversitesinde son sınıf ELT öğrencileri ve bir ELT profesöründen toplanan 12 saatlik videoya kaydedilmiş sınıf etkileşim verilerine dayanmaktadır. İlaveten bu çalışma, test türleri ile öğretmen adaylarının yönelim gösterdiği ölçme ve değerlendirme ilkeleri arasındaki ilişkiyi

ortaya çıkarmaktadır. Yapılan analiz göstermiştir ki, akran geribildirim etkileşimi sırasında ölçme ve değerlendirme ilkelerine başvurmanın dil öğretmenleri için gerekli olan yabancı dilde ölçme ve değerlendirme okuryazarlığını ve becerilerini geliştirmek için öğrenme fırsatları sağlamaktadır. Elde edilen sonuçlara göre, bu çalışma yabancı dilde öğretmen eğitimi ve ölçme ve değerlendirme okuryazarlığının yüksek öğrenim düzeyinde geliştirilmesi için çıkarımlar ve öneriler sunmaktadır.

Anahtar Kelimeler: Söylem Çözümlemesi, Sınıf İçi Akran Etkileşimi, Yabancı Dilde Ölçme ve Değerlendirme, Ölçme ve Değerlendirme İlkelerine Başvurma, Yabancı Dilde Öğretmen Eğitimi

To my family…

ACKNOWLEDGMENTS

First and foremost, I would like to extend my gratitude to my thesis supervisor Prof. Dr. Çiler Hatipoğlu, who shed light on the research journey that I have been following for the past three years. Without her inspiration and support, I would not complete this thesis. Since the day we have met, she has become a source of inspiration for me and encouraged me to go further in my career. I am one of the luckiest graduate students to find a chance to work with her.

Moreover, I would like to express my deep gratitude to Asst. Prof. Dr. Nilüfer Can Daşkın for her valuable time and contributions on my thesis. Thanks to her constructive feedback and valuable insights on this thesis, I have been able to find my way. I should also thank Assoc. Prof. Dr. Perihan Savaş for her time and guidance on this thesis.

I also thank Assoc. Prof. Dr. Ufuk Balaman, for sparking the CA interest in me since my undergraduate years.

I would also like to thank Mehmet Gazan for supporting me in every step I take. He has been the most valuable companion in my life. My warmest thanks go to my best friend Pınar Topal whom I have stepped into various academic journeys since Hacettepe. I also appreciate Cennet Çalışmış for being my side as my teacher-bestie since 2017. I am also grateful to my long-distance best friend Julien Moysan for taking a part on my never-ending academic journey.

My deepest gratitude is for my family. Their love, trust and faith in me always helped me go further. Lastly, I should thank my students for making each day in my life brighter and warmer.

TABLE OF CONTENTS

PLAGIARISM....iii
ABSTRACT....iv
ÖZ....vi
DEDICATION....viii
ACKNOWLEDGMENTS....ix
TABLE OF CONTENTS....x
LIST OF TABLES....xii
LIST OF FIGURES....xiii
LIST OF ABBREVIATIONS....xiv
CHAPTERS
1. INTRODUCTION....1
1.1. Statement of the Problem....1
1.2. Aim and Significance of the Study....3
1.3. Assumptions....6
1.4. Limitations....6
1.5. Definitions....7
2. LITERATURE REVIEW....8
2.1. Interaction in Preservice Teacher Education....8
2.2. Language Testing and Evaluation in Preservice Teacher Education....15
2.2.1. Language Assessment Literacy (LAL)....17
2.2.2. English Language Testing and Evaluation (ELTE) Course....18
2.3. Peer Feedback as Formative Assessment....25
2.3.1 Peer Feedback and Preservice Teacher Education....29
2.3.2. Perceptions towards the Implementation of Peer Feedback in Higher Education Context....31
3. METHODOLOGY....34
3.1. Conversation Analysis....34
3.2. Transcription, Building a Collection, and Data Analysis....37
3.2.1. Reliability....40

3.2.2. Validity 41
3.3. Participants and Research Context 43
4. ANALYSIS AND FINDINGS 48
4.1. Employment of RTP in Problematization Phase 48
4.2. Employment of RTP in Suggestion Phase 76
4.3. Emergence of RTP in Resistance & Resistance Management Phase. 94
5. DISCUSSION 117
5.1. Emerging Recurrent Phases in the Peer Feedback Interaction 118
5.1.1. Problematization 118
5.1.2 Resistance 121
5.1.3. Suggesting a Change 124
5.2. Functions Performed by RTP 126
5.3. Testing Principles Oriented in Three Recurrent Phases of Peer Feedback Interaction 128
5.4. Learning Opportunities RTP Offers throughout the Peer Feedback Sessions 132
5.4.1. Preservice Teachers' Display of Uptake in Follow-Up Turns 132
5.4.2. Preservice Teachers' Display of Uptake in Self-Feedback Turns 133
5.4.3. Teacher's Invoking Larger Pedagogical Principles in Her Follow-Up Turns 133
5.5. Implications 135
5.5.1. Implications For L2 Teacher Education and ELTE Course 135
6. CONCLUSION 138
6.1. Limitations 139
6.2. Suggestions for Further Research 139
REFERENCES 141
APPENDICES
A. JEFFERSON TRANSCRIPTION CONVENTON (2004) 168
B. MONDADA MULTIMODAL TRANSCRIPTION CONVENTION (2018) 169
C. ETHICS COMMITTEE APPROVAL 170
D. TURKISH SUMMARY / TÜRKÇE ÖZET 171
E. THESIS PERMISSION FORM / TEZ İZİN FORMU 183

LIST OF TABLES

Table 1 Number of 'RTP' in Feedback Sessions 40
Table 2 ELTE Course Outline 45
Table 3 Peer Feedback Groups 46
Table 4 The Duration of Recordings 47
Table 5 Oriented Testing Principles in Three Sequential Environments 131

LIST OF FIGURES

Figure 1 Example of the Phenomenon (RTP) 5
Figure 2 Sequential Environments RTP Emerges in the Peer Feedback Interaction 118

LIST OF ABBREVIATIONS

CA	Conversation Analysis
CIC	Classroom Interactional Competence
ELT	English Language Teaching
ELTE	English Language Testing and Evaluation
FA	Formative Assessment
FLT	Foreign Language Teaching
FPP	First Pair Part
HE	Higher Education
L2	Second/Foreign Language
LAL	Language Assessment Literacy
PF	Peer Feedback
RTP	Reference to Testing Principles
SPP	Second Pair Part
TCU	Turn-Constructional Unit
TRP	Transition Relevance Place

CHAPTER 1

INTRODUCTION

This chapter briefly introduces the statement of the problem and continues with the aim and significance of the study. Next, the research questions guiding the study and the terminology are presented. The chapter is concluded with assumptions and limitations of the study.

1.1. Statement of the Problem

Second language assessment, described as a "notoriously difficult domain of knowledge" (O'Loughlin, 2006), puts a significant amount of pressure on prospective teachers to be assessment literate to conduct *sound assessment practices* in their classrooms (Stiggins, 1995). However, as Stiggins (1999) states, "many teachers are left unprepared to assess student learning as a result of insufficient preservice and in-service training." In this regard, a majority of research has indicated that language teachers have low levels of L2 assessment literacy, which negatively affects the quality of instruction and hence teaching.

In her study examining the scope and content of language testing and evaluation (ELTE, hereby) courses in Turkey, Şahin (2019) has demonstrated that a single ELTE course is not enough to include all the practical and theoretical concepts necessary to construct the language assessment knowledge and literacy of prospective language teachers during their graduate studies. According to the earlier research, ELTE courses generally place a huge emphasis on the summative aspect of testing by focusing on teaching theoretical concepts (terminology, test types, testing language skills, etc.) which results in the dominant use of traditional testing methods and negligence of using alternative assessment methods such as self-feedback, peer-feedback and so

forth. The studies clearly show that the learning objectives of ELTE courses in international and Turkish higher education contexts fail at leading preservice teachers towards putting their theoretical knowledge into implementing assessment tasks in their future careers. In line with this, several studies have been carried out to understand the beliefs and needs of preservice and in-service teachers on assessment to structure the syllabus of ELTE courses in a better way. According to the results, it has been concluded that preservice teachers are in dire need of further hands-on training on assessment.

Despite this, the studies conducted on understanding the training needs of preservice teachers have paid far too little attention to the actual learning processes in the classroom. According to Johnson (1999) and Shulman (1987), “teachers' knowledge is constructed through experiences in and with students, parents, colleagues, and administrators; we can say that the process of learning to teach is socially negotiated.” In this regard, a considerable gap in teacher education research needs to be filled in terms of investigating how teachers learn by socially interacting with their close surroundings in the classroom. In other words, preservice teachers' classroom interaction is a sine qua non of teacher education studies concerning sociocultural perspective.

In this respect, the earlier studies (Bachman, 2018; Norries et al., 1998) on language assessment & evaluation mainly focused on addressing the teacher's needs, beliefs, and perceptions towards teaching subject matters, e.g., L2 assessment. On the other hand, interactional studies (Akbari, 2007; Carlo, 2010; Matthew, 2017) mostly center upon reflective practices such as observation, feedback sessions, and dialogic talks that aim to enhance teaching quality by holding teacher/student discussions. Therefore, a significant amount of literature in teacher education focused on the question of 'how to teach. However, far too little attention has been paid to the question of ‘how to test’ In this respect, the existing literature fails to bring an account for how preservice teachers learn through their interactions with their peers and what is actually going on during this learning process in the classroom. Therefore, to fully grasp the learning process teachers go through in their development of L2 assessment skills and

knowledge in ELTE courses, microanalytic research investigating the interactional practices of preservice teachers is needed.

1.2. Aim and Significance of the Study

This study conducts a micro-analytic investigation into the preservice teacher interaction in the English language testing and evaluation course (ELTE) in the department of foreign language education at a state university in Turkey. The study aims to develop an understanding of how preservice teachers develop their L2 assessment skills and literacy, which form a crucial part of their pedagogical content knowledge (Shulman, 1987) through their social interactions with one another.

According to Freeman and Richards (1993), "what teachers know about teaching is largely socially constructed out of the experiences and classrooms from which teachers have come." Therefore, looking into the interaction of preservice teachers within the context of the L2 assessment and evaluation course contributes to the literature in teacher education research in the following ways:

To begin with, in their famous work that reconceptualizes the teacher education framework Freeman and Johnson (1998) label teachers as 'learners of teaching' and "central to understanding and improving English language teaching." In parallel with this, this study aims to gain a deeper understanding of the learning process preservice teachers go through by closely observing and analyzing their talk-in-interaction in situ, which "becomes a vehicle for understanding" (Hall, 2003) the development of preservice teachers' classroom interactional competence (Walsh, 2011).

Secondly, examining the interactional practices of preservice teachers in ELTE courses through the microanalytic perspective of conversation analysis provides a better understanding of how preservice teachers acquire L2 assessment skills and improve their assessment literacy through peer feedback interactions. As Stiggins (1995) states, a well-educated and assessment literate teacher is someone "who knows the difference between sound and unsound assessment" is a pre-requisite to meet the changing needs of assessment and evaluation in various educational contexts (Hatipoğlu, 2015).

Lastly, as the peer feedback interactions of preservice teachers set the context of the study, this thesis also illustrates how to integrate peer feedback into the curriculum of preservice teacher education programs and what sort of benefits can be drawn for further use of training practices which lag behind the modern approaches to L2 teacher education.

In this respect, previously conducted research in teacher education studies mainly emphasized two research foci: 1) teacher cognition and defining teacher's beliefs & perceptions 2) teacher reflection practices in the form of written journals and feedback sessions through the medium of different methodologies such as action research, narratives, case studies, etc. Similarly, the literature on language testing and assessment centered upon the assessment literacy of preservice and in-service teachers, their needs, and beliefs, and lastly the content and organization of ELTE courses in teacher education programs. However, despite a bulk of research conducted through various methodologies, interaction studies that provide a data-driven and insider-account approach to teacher education have been quite limited. In this regard, according to Hale et al. (2018), "one of the challenges facing both teacher educators and practitioners is to identify, formulate, and share tools that promote dialogically, engaged, and evidence-based practice." As a result, conversation analysis has been adapted as the research methodology of this thesis and therefore the study differs from the previous literature by basing the focus on the interactional practices of preservice teachers in a teacher education context.

Drawing on the conversation analytic approach, the study is based on the peer feedback interaction data gathered in language testing and evaluation course in an English language teaching (ELT) context. It has uncovered the phenomenon "Reference to Testing Principles" (RTP) which emerges in different sequential environments such as 1) problematization, 2) resistance 3) suggestion. To exemplify the phenomenon under investigation, an extract from the data set (Figure 1) has been provided below.

(Week 6)- 7.45 / 9.23 Commenting on the reading section, *superhereos* of group 1.

(Uze raises his hand and T nods her head)

1 **UZE:** **hocam correct me if ı am wrong right *ı once read heaton page one hundred**
2 ***four(0.2) or five ı dont remember**

>>*---1--->*0.7

1:bends his head and merges his palms on his face

3 **T:** **°hm mm°**
4 **UZE:** **but he sa::ys() if you have scanning questions question should come**
5 **[*first***
>>*--2--*>
UZE: 2:moves his right hand upperwards

6 **İLK:** **[yeah**

7 **UZE:** **then the text should come er later beca:use the questions are going to**
8 **lead the students to find the information without you know mhm[spending**
9 **too much time=**

10 **T:** **=exactly**

Figure 1 Example of the Phenomenon (RTP)

In the extract above, group 2, consisting of ELA, PER, MER, FEY, SOR in the 6th session of the course, receives feedback on the reading sections of their language exams which consist of scanning questions in the form of multiple-choice questions. However, a member from the feedback-provider group, UZE, self-selects and starts a problematization sequence for instruction placement in the scanning questions in line 1. During the formulation of his problematization UZE gives a reference to a famous testing book, Writing English Language Tests (1990), and its author, J.B. Heaton, whom he has studied from the assigned readings list in the course. He problematizes the misplacement of the instruction, which must come before the question stem to lead the students to find the expected information in the text. While constructing his turn, UZE brings an account of his problematization from a famous course book and author before his problematization as evidence to support his claim and performs a self-policing. His account includes the phenomenon of this study, invoking testing principles, which is "instructions should be placed before the question stem." It can be concluded that preservice teachers back up their claims with testing principles which

they resort to during their feedback delivery to the item-writer groups. Therefore, this study aims to bring an in-depth analysis of the phenomenon of 'reference to testing principles' during the peer feedback interaction sessions of preservice teachers in an English language testing and evaluation course at a state university in Ankara, Turkey drawing on the analytic approach of conversation analysis.

Therefore, in line with the moment-by-moment analysis of preservice teachers' interactional practices *in situ*, the study seeks to answer the following research questions:

1) What are the sequential environments does the RTP emerge in peer feedback interaction?

2) What are the main functions RTP perform in peer feedback interaction?

3) What are the principles oriented by preservice teachers in different sequential environments?

4) How does RTP provide preservice teachers with learning opportunities for their assessment skills?

1.3. Assumptions

It is assumed by the researcher of this study that the conversation analysis methodology fits the examination of the naturally occurring data set since the teacher education programs mostly suffer from engaging their trainees/student teachers in the use of tools that is based on authentic, evidence-based, and detailed observations.

1.4. Limitations

The limitation of this study results from the limited participants in the ELTE course which took 6 weeks during the summer school period of the academic year 2018 in the department of foreign language education at a state university in Ankara, Turkey. Although the data set is composed of 6 successive weeks of an intense module ELTE

course, it already consists of 12 hours of video recordings which are assumed to be quite enough for conducting a microanalytic investigation (Seedhouse, 2004a).

1.5. Definitions

Pre-service teachers: Undergraduate students enrolled in English language teaching departments who train to become certified English language teachers on the condition of completing their 4-year graduate program consisting of pedagogical content courses and teaching practicum.

ELTE: English language testing and evaluation course which is a core course in the curriculum of English language teaching graduate programs. The course entered the teacher education curricula in 1998 by the higher education council and ministry of education to equip teachers with the necessary technical and practical knowledge of assessment in L2 (Hatipoğlu, 2017).

Peer Feedback: A type of formative assessment through which students/learners actively engage in the collaborative task of providing feedback on each other's performance or work such as a project, assignment, etc. concerning its quality, correctness, and so forth. The process offers opportunities for students to learn from one another and become more autonomous in their learning process (Falchikov, 2006).

Testing and Assessment Principles: A set of rules providing a guideline to ensure the construction of valid, reliable, practical, and effective language tests (Heaton, 1990, p. 114).

CHAPTER 2

LITERATURE REVIEW

This chapter outlines a review of the existing literature. The first section mainly places emphasis on social interaction in relation to teacher education studies. In the second section, studies conducted on second language assessment and testing are discussed. Thirdly, a review of studies in peer feedback is presented.

2.1. Interaction in Preservice Teacher Education

"In many respects, the fundamental or primordial scene of social life is that of direct interaction between members of a social species, typically ones who are physically co-present*."* (Schegloff, 1996a as cited in Fitch & Sanders, 2005, p. 87).

Amongst a myriad of approaches to understanding the nature of learning and teaching, studies on interaction have gained prominence as "a vehicle" to discover learners'and teachers' meaning-making and knowledge-building processes. In parallel with this, studies on interaction vary from classroom interaction context to institutional context, and several other research settings.

In this respect, interaction in preservice teacher education has also gained its place as a distinct field of study in recent years within the sociocultural paradigm shift in second language education (Jacobs et al., 2001). When the existing literature is reviewed, a great number of researchers (Hale et al., 2018; Markee, 2000, Sert & Seedhouse, 2011; Walsh, 2001, 2013) stand out with their specific research foci to understand the mechanisms underlying teacher education programs. In this regard, the key aspects of interaction studies conducted in the teacher education context are described as follows:

To begin with, one of the current hot topics in the teacher education research context since the 1970s has been the study of teacher cognition. Previous studies on teacher cognition which investigate "how teachers learn to do their work" (Johnson, 1994) focused on individual factors such as "schooling background, professional experience, and classroom practice" (emphasis in original) that tailor teachers behavior and instructional decisions throughout their career" (Borg, 2003). However, as Johnson (2009) states, "since teachers' knowledge of teaching is constructed through experiences in and with students, parents, colleagues, and administrators, the processes of learning to teach are socially negotiated". In agreement with this social aspect of cognition, Li (2017, 2020) coined the term *cognition-in interaction* which provided the concept of teacher cognition with a more *discursive* stance (Sert, 2019), and emphasized its constant development through the dynamic interactions in the specific social context in tandem with teacher's beliefs, attitudes, and perceptions.

Additionally, the relationship between teachers' beliefs, perceptions, and their reflection on teaching is another dominant research field in teacher cognition (Fang, 1996; Kagan, 1988). In this regard, several studies have addressed how teachers' beliefs unfold in their instructional decisions in general teaching (Borg, 2003, 2006; Li, 2012, 2020; Sato & Kleinsasser, 2004) and in teaching specific language skills such as grammar, reading, and writing. The results demonstrate that teachers' have a "filter of beliefs" (emphasis in original) through which their decisions are shaped during their teaching of the subject matter (Andrew, 2003; Johnson, 1992; Shavelson, 1983). However, previous research findings into the effect of teachers' beliefs have been inconsistent and contradictory. In this respect, while some studies have found a consistent relationship between teachers' beliefs and their teaching practice (Erkmen, 2014; Kızılay, 2018; Kuşçu, 2014; Saraç-Süzer, 2007) other studies highlighted the contextual factors such as classroom atmosphere and learners' attitudes & needs on shaping teachers' behavior in their instructional practices (Baştürkmen, 2012; Çalışır-Gerem & Yangın Ekşi, 2019; Tüllüce, 2019).

Yet, to make sense of teachers' day-to-day instructional practices, researchers in the last two decades, have focused on prospective methods such as reflective practices in preservice and in-service teacher education research instead of dwelling on

quantitative studies of retrospective past experiences or context-bound belief systems that teachers possess (Merryfield, 2009).

In this regard, teachers can make use of a variety of reflection practices such as self-assessment, writing journals, or peer assessment method via dialogic talks or classroom observations through mentor-guided post-observation sessions. Since the characteristics of the teaching context continuously undergo changes every year or semester, reflective practices can be very fruitful when preservice teachers are offered the chance to use their hidden potential to choose the most suitable way of providing reflection and hence contribute to the teacher's professional development or growth (Demirbulak, 2012).

In this respect, a wide range of studies conducted on reflection practices through the employment of different methodologies such as reflective journals (Korkmazgil, 2020; Mathew, et al., 2017) action research (Dinkelman, 2000; Güngör, 2016, Zeichner, 2006) grounded theory (Douglas, 2003; Rodman, 2010) case study (Farrel, 2012; Goodman, 1984) as well as conversation analysis (Ghafarpour, 2016; Hale et al., 2018; Lazaraton & Ishihara, 2005; Li & Walsh, 2011, Walsh & Mann, 2015).

What these studies have found in common is that when teachers or peers engage in reflection practices, they become “each other’s eyes” (Brookfield, 2017) and easily pinpoint the weaknesses and strengths in the specific classroom interactional context to enhance the quality of teaching performance and student learning. Moreover, concerning reflective practices, interest in interaction-based teacher education models aiming at “long-term professional development” *(*Copland et al., 2009) grows continuously among researchers (Sert, 2015; Walsh, 2006, 2011; Waring, 2020). As Crandall (2000) states, “there is a growing sense that language teacher education programs have failed to prepare teachers for the realities of the classroom”. In line with this statement, *data-led* teacher education models are developed to offer more *authentic, evidence-based reflection* practices in L2 teacher education programs. (Walsh & Mann, 2015). To exemplify, Walsh’s (2001, 2003) teacher education model SETT (Self-evaluation of teacher-talk) developed as an *ad hoc approach* to direct teachers to *self-observation* and *self-reflection* of their instructional practices.

The framework was born out of the need for equipping teachers with more detailed and immediate feedback within the analytical framework of CA. Hobbs (2007), Walsh (2006), and Walsh and Mann (2015) criticize teacher reflective practices which are heavily based on written forms in terms of being "mechanical, generic, inauthentic" (emphasis in original) and too concerned with assessment and evaluation rather than providing reflection. In this respect, SETT framework guides teachers through identifying facilitative and obstructive learning practices by focusing on the specific micro-learning context and pedagogical goal of the lesson via actual recordings of classroom interaction (Walsh, 2003). Ultimately, teachers pay direct attention to the task of critically observing and evaluating the lesson plans and instructional practices which contribute to their pedagogical knowledge and raise their awareness of the use of language.

In line with the aim of SETT another framework, SWEAR, is recently developed by Waring (2020). The purpose of the framework is similar to SETT in terms of raising teachers' awareness in locating instructional problems, participating in discussions, and providing solutions in cooperation (Sert, 2019).

In addition to these, Sert (2015, 2020a) has presented another reflective teacher education framework called IMDAT which consists of 3 main steps as follows: teaching, reflection, and feedback. Sert (2015) has introduced the concept of Classroom Interactional Competence (CIC, henceforth) developed by Walsh (2006, 2011) which is defined as "the ability to use interaction as a tool for mediating and assisting learning" into the reflection practices of in-service teachers through the medium of classroom interaction analyses (p. 56). According to Sert (2015) and Walsh (2011), when the literature is reviewed, teacher mentors traditionally observe the teaching performances of pre-service and in-service teachers by doing classroom visits and holding observation reports and writing on checklists which only offers "mere coaching" and "mere grading" to novice teachers yet fail at passing on their experiences and stimulating opportunities for reflection.

In parallel with this, directing teachers to critically evaluate their teaching performance based on classroom interactional data helps them in becoming more self-conscious and self-monitoring in their teaching context.

Suffice to say, teacher education frameworks are proven to be highly effective and beneficial for the development of teachers' L2 CIC and language awareness when they are fully utilized by actual recordings of classroom interactions in feedback sessions and dialogic reflections in tandem with journals and interviews if considered necessary (Sert, 2019).

Besides this type of research conducted on teacher education frameworks, post-observation feedback conferences and the interaction dynamics between teacher mentors and teacher trainees have also caught great attention from researchers (Copland, 2009; Engin, 2014; Gwyn-Paquette & Tochon, 2002; Kim & Silver, 2016; Vazquez, 2009; Waring, 2013, 2017). The researchers highlighted that besides the undeniable effect of post-observation conferences on novice teachers' improvement in teaching, the studies are important in terms of shedding light on the nature of mentor-trainee talk; how it unfolds, and how it fosters or impedes reflection when factors such as social identity, context, relationship are taken into consideration. In this respect, while the mentor and trainee relationship suffer from a variety of challenges, the role of interaction in teacher education framework remains as an important field which is newly discovered (Hale et al., 2018, 2022) especially in Turkish context (Sert, 2011, 2013).

As Mann and Walsh, (2015) highlight "it has been claimed that observation practices have to drive reflective practice developed through interaction". In this regard, a limited number of research on peer feedback interaction studies in teacher education research also gained momentum in recent years. International studies in this field, (Battle & Seedhouse, 2020; Philips, 1999; Strong & Baron, 2004) generally draw on how the peer feedback interaction is distinctively put into practice and what kind of expected outcomes are described at the end of the process. On the other hand, mixed-method studies conducted in the Turkish context (Göker, 2006; Koç & Ilya, 2016; Yüksel, 2011; Yüksel & Başaran, 2020) investigated the possible drawbacks and

benefits of peer feedback interaction. The studies have revealed that while personal relationships might hinder the implementation of peer feedback in the right way, peer feedback is still an effective tool in the promotion of collaborative reflection and enhancement of critical thinking skills in teacher education research.

Suffice to say, when the relevant literature in terms of interactional studies in the teacher education context is carefully examined, it has been shown that a wide range of studies (Atkinson, 2000; Brookfield, 1995; Johnson, 1996) have been undertaken on reflective practices and feedback sessions at the tertiary level. In addition, the studies mainly focused on dialogic talks, post-observation conferences, and peer feedback sessions in the L2 teacher domain. So far, however, there has been little discussion about classroom interaction research in relation to teacher education. In this regard, interaction studies are important in terms of three main aspects: First as Battle and Seedhouse (2020) states, "the absence of data-led analysis*"* clearly pinpoints to the problem of leaving the teachers outside of the reflective practices, and "the lack of juxtaposition between teachers' perceptions and their actual practices in situ" (Ghafarpour, 2016). There is a great need in teacher education research for observing the learning processes emerging in teachers' social conducts which can only be met by adopting an analytic approach (Walsh & Mann, 2015) that is, conversation analysis. To fully grasp the nature of teacher learning and development of disciplinary knowledge, one must study the social processes teachers go through in their cognitive states in and through interaction (Doehler & Lauzon, 2015). According to Seedhouse and Walsh (2010), "learning is defined as a change in cognitive states." But how one can understand this change? Seedhouse (2010, p. 127) conceptualizes the learning through a period as follows:

1) The first phase involves the gap regarding the use of target item such as *a lexical item, or pattern* in learners' state which shows that the learner cannot perform a *micro feature* of a language and needs scaffolding.

2) The second phase takes place when the learner starts performing the target item in a similar context but without the scaffolding.

3)The third phase describes the learners' independent re-use of the target item in a new environment which brings evidence to the change in his/her cognitive state.

All things considered, I will refer to Seedhouse' s conceptualization of learning in terms of the change in learner's cognitive state in the context of preservice teachers in language assessment and evaluation course on the analysis chapter of the study.

Returning to the subject, as it is exemplified above interactional studies can clearly portray teachers' continuous and complex learning process id est, "learning-in action" (Ellis et al., 2010) through the micro lenses of conversation analysis in the most detailed manner (Markee, 2015). Lastly, they can illustrate how novice and experienced teachers develop their classroom interactional competence (CIC) and how it affects the quality of learning and teaching.

In view of all that has been mentioned so far, teacher-cognition and teaching practice skills, thereby teacher professional development, have become the research foci of interactional studies conducted in the teacher education context in the last three decades (Freeman & Johnson, 1998). In parallel with this, one may conclude that an increasing number of researchers in teacher education context have sought answers to the famous question of "how to teach*?"* while "how to test?*"* often remains as terra incognita within the domain of teacher pedagogical content knowledge (Shulman, 1987) In this respect, among the 6 categories that define teachers' knowledge base, Shulman (1987) labeled the content knowledge as the most important 'province' of teachers. Therefore, to be able to fully understand the process teachers go through in their acquisition of pedagogical content knowledge and development of skills in L2 testing and assessment, moment-by-moment analysis of classroom interaction in undergraduate English language evaluation and assessment courses is of importance. In this respect, the question of "how teachers commute from the status of the learner to that of teacher" (Shulman, 1987) can only be answered if research deeply investigates the interactional practices of preservice teachers in their unique classroom context to be able to inform the teacher education programs about the current needs of preservice teachers to make necessary changes.

2.2. Language Testing and Evaluation in Preservice Teacher Education

Since assessment is a crucial element in today's education system, there is a growing demand for teachers to assess the expected learning outcomes in their teaching context to revise the program, adjust the curriculum, and, most importantly, promote learning efficiently (Brindley, 1998). As Crooks (1988) and Stiggins (1999) notes, "the typical teacher can spend as much a third to a half of his or her time involved in assessment-related activities." Therefore, preservice teachers need to acquire the necessary assessment skills to understand whether the learning process will result in a favorable outcome or failure (Davies, 2013). The systematic review of the literature demonstrates that assessment needs & literacy levels of preservice and in-service teachers, their beliefs, and attitudes in L2 assessment, as well as analysis of ELTE courses, are the focus of an extensive range of previous studies. In this respect, a considerable amount of literature has been published on understanding the assessment needs of teachers in international context (Cheng, Rogers & Hu, 2004; Frodden, Restrepo & Maturana, 2004; Fulcher, 2012; Hasselgreen, Carlsen & Helness, 2004; Munoz, Palacio & Escobar, 2018; Sheehan & Munro, 2017; Vogt, Tsagari & Spanoudis, 2014). Besides research conducted in the Turkish context (Işık, 2020; Kavaklı & Arslan, 2019; Köksal, 2014, Ölmezer-Öztürk & Aydın, 2018) is in complete agreement with the international community on providing more training in assessment for L2 teachers. In that matter, findings suggest that assessment-related needs and expectations of teachers should be fulfilled during their undergraduate courses. To address the issue more comprehensively, the studies published by Köksal (2014) and O'Loughlin (2013) suggest that providing more courses on assessment through online tutoring sessions and workshops for language testers, in general, is crucial for the attainment of assessment competency. However, Yan, Zhang and Fan (2018) highlight the factors behind the needs of teachers as L2 teachers' profile, their teaching context, and their perceptions and attitudes towards assessment. In this respect, the design of evaluation and assessment courses and training programs must consider these factors (Vogt, Tsagari & Spanoudis, 2020).

As a matter-of-fact teachers' beliefs and attitudes shape their approach to teaching, learning, and hence assessment (Cizek et al., 1996). In parallel with these, numerous

studies have examined teachers' beliefs, attitudes, and perceptions concerning their effect on classroom practices. However, there are contradictory findings obtained from the previous studies. Much of the current literature in Turkey (Kavaklı & Arslan, 2019; Ölmezer-Öztürk & Aydın, 2019) has revealed that teachers' lack of knowledge and competence in assessment is in line with their negative attitudes. On the other hand, studies conducted by (Black & William, 2005; Brown, 2004; Cheng, Hu & Rogers, 2004; Munoz, Palacio & Escobar, 2012; Roger et al., 2017; Shohamy et al., 2008) and also (Büyükkarcı, 2014; Gonen & Akbaraov, 2015; Karagül, Yüksel & Altay, 2017; Öz & Atay, 2017) in Turkish setting demonstrate that although teachers have positive attitudes on assessment, there is a gap between their beliefs and actual classroom practices dwelling on summative evaluation methods heavily. According to Yorke (2005), implementation of formative assessment practices such as self & peer feedback and portfolio in language classrooms are among the hot topics being discussed within the last three decades. Yet, there is still a strong tendency to utilize summative evaluation methods at all levels of education. A major contributing factor for choosing commonly preferred ready-made tests (Merter & Campbell, 2005; Şişman & Büyükkarcı, 2019), which mainly consist of question types such as multiple-choice items and fill-in-the-blanks is the standards-based education system (Cheng, Rogers & Hu, 2004). Since the placement of students into secondary and tertiary level schools is generally based on scores obtained from national university placement exams, teachers and learners pay particular attention to the use of specific test types, which are designed in the traditional method of testing; summative assessment (Llosa, 2011). Other factors affecting language teachers' choices are personal and institutional constraints (Burns, 1996) resulting from overcrowded classrooms, short course hours, and lack of training in alternative assessment methods. Within this context, the commitment to traditional testing methods will inevitably result in the fossilization of assessment skills and the negligence of the learning aspect of teaching. Consequently, the need for a 'student-centered' alternative approach to testing is vital for training L2 language teachers and increasing their assessment literacy to benefit learning.

2.2.1. Language Assessment Literacy (LAL)

In this sense, knowledge in assessment lies at the heart of language assessment literacy (LAL, hereafter) (Xu & Brown, 2017). Inbar-Lourie (2012) defines LAL as "refers to the knowledge skills and principles that stakeholders involved in assessment activities are required to master to perform assessment tasks in the classroom." The sociological shift toward a formative assessment approach in assessment has highlighted the priority for the acquisition of assessment knowledge; therefore, increasing number of research between the period 1991- the 2000s have examined the LAL levels and needs for preservice (Giraldo & Murcia, 2018; Wang, Wang & Huang, 2014); and in-service teachers (Giraldo, 2017; Mertler, 2003; Schaffer, 1993; Stiggins, 1995; Taylor, 2009; Weng & Shen, 2022). The findings of the studies demonstrated that language teachers' assessment level is insufficient, and therefore teachers lack assessment skills and sound knowledge to conduct assessment-related activities in their careers (Tsagari & Vogt, 2014; Xu & Brown, 2017).

In Turkish context recent studies conducted by (Büyükkarcı, 2016; Genç, Çalışkan & Yüksel, 2020; Hatipoğlu, 2015b; Mede & Atay, 2017; Ölmezer-Öztürk & Aydın, 2018; Şişman & Büyükkarcı, 2019; Valizadeh, 2019; Yeşilçınar & Kartal, 2019) have investigated LAL levels of in-service language teachers while (Kavaklı & Arslan, 2019) focused on teacher candidates. The findings of the studies align with previous research, thereby revealing that preservice and in-service language teachers in Turkey are in dire need of developing their levels of LAL. A more comprehensive review study was carried out by Sevimel Şahin & Subaşı (2019) to compare the studies conducted in LAL in the Turkish context and EFL contexts around the world. The results demonstrated that undergraduate teacher training is insufficient in general course structure in relation to the integration of assessment theory and practice. Therefore, language teachers from preservice to in-service have poor language assessment literacy.

Another research by Yeşilçınar & Kartal (2020) has indicated a mismatch between the dual role of instructor and assessor. According to the study, language teachers did not assume the assessor's identity. The study suggests that quality teacher training is a vital

factor in undertaking the role of an assessor from being a novice teacher. In this respect (Işık, 2021; Öztürk & Aydın, 2019; Xu & Brown, 2017; Yastıbaş & Takkaç, 2018) have focused on other factors affecting the development of language teachers' LAL and listed them as an academic degree, experience, teaching context, culture in assessment and policies in education. Büyükkarcı (2016) also added that training after graduation and in-service experience do not endow teachers with a satisfactory level of LAL to perform sound assessment activities in their careers.

Thus far, several studies have linked the lack of adequate teacher training with low levels of language assessment literacy. Therefore, language teacher education programs lay the foundation for the attainment of LAL; that is, principles, knowledge, and skills necessary for being a qualified language teacher who can conduct sound assessment tasks in class and announce the result to provide feedback for their students to enhance the learning outcomes.

In conclusion, equipping language teachers with LAL to foster the implementation of formative assessment methods and promote teaching and learning is crucial (Davison & Leung, 2009). In this respect, although much of the research reports findings on teachers' LAL and their reflection on classroom practices, it is necessary to draw attention to the preservice teachers' training needs, especially in undergraduate assessment and evaluation courses.

2.2.2. English Language Testing and Evaluation (ELTE) Course

One of the teacher's primary duties of a teacher is to assess students' performance periodically and provide feedback for further action (Merter & Campbell, 2005). In this sense, while teachers dedicate “almost half of their professional time to assessment-related activities” (Stiggins, 1993), they might be held accountable for the success of their students, program, and even the institution they work at (Hatipoğlu, 2010). In parallel with this, language assessment, testing, and evaluation courses have gained prominence in preparing preservice teachers to acquire the necessary assessment skills for their prospective careers. Thereby ELTE has become one of the compulsory courses in the curriculum of teacher education programs worldwide

(Brown & Bailey, 1999). In this regard, there is a bulk of research on teachers' assessment literacy, needs, and beliefs. However, language testing and evaluation courses (ELTE) haven't received enough attention. (Brown & Bailey, 1999; Hatipoğlu, 2010; Inbar-Lourie, 2008; Jin, 2010; Johnson et al. 1999; Kleinsasser, 2005).

In this respect, several researchers stand out with their research on the content of assessment courses (O' Loughlin, 2006), their efficacy (Johnson et al., 1999), and students' attitudes (Giraldo & Murcia, 2019) towards them. Among these researchers, the studies conducted by Brown& Bailey (1999, 2008) and Jin (2010) have investigated the course content of ELTE along with teachers' beliefs and students' perspectives towards the course.

To illustrate, Brown and Bailey (2008) re-examined the L2 testing and evaluation courses in terms of content, effectiveness, and teachers' background as well as students' attitudes in a quantitative-designed study as a complementary work to their research in 1999. Compared with courses in 1999, topics with the highest mean rating were hands-on experience, test critiquing, and test analysis, while administration of tests and test-taking had the lowest mean ratings. In terms of attitudes of undergraduate and graduate students towards the course content, the data revealed being" satisfied, less scared, and confident was among the popular answers given to the interview questions at the end of the course. Therefore, the study shows an intersection between the studies conducted in 1999 and 2008. Yet, it offers solid suggestions on integrating practice and theory in L2 assessment & evaluation courses and developing students' assessment skills through first-hand experience to increase the quality of teacher education programs in the 21st century.

Another study by Jin (2010) investigated the course content, methodology, and perceptions of students in ELTE courses in China. Preliminary findings reported that Chinese universities do not differentiate in their course content to a great extent. They successfully cover the main theoretical concepts in assessment. Yet, because of insufficient hours, courses generally ignore the inclusion of topics such as classroom practice which includes item writing & reliability analysis and educational-psychological measurement id est statistics. Although students' perceptions of the

courses were positive, L2 testing, and evaluation courses designed in traditional assessment format are beyond providing teachers with what they should know (Stiggins, 1991b).

In this sense, the study critiques L2 testing and evaluation courses in China. It offers suggestions to increase the course hours and reform the course content so that students can put their theoretical knowledge into classroom practice to improve the standards in L2 language teachers' education programs.

While these studies focused on the layout of ELTE courses and teachers' points of view in their implementation, they did not consider the student's needs, beliefs, and perceptions. From this point of view, Kleinsasser (2005) and Hatipoğlu (2010) addressed the students' needs and conveyed the findings to develop the scope of ELTE courses and the quality of teacher education programs.

Firstly, Kleinsasser (2005) transitioned from teacher-focused to a learner-centered model in MA, TESOL program on language and testing course in collaboration with postgraduate students. The course's primary objectives were to engage students in critically reviewing assessment literature and analyzing the tests based on format, design, appropriacy, and validity in class discussions. However, in the phase of criticizing instruments, students are involved in assessment-related practices in class and developing original assessment materials. In that vein, the study set an excellent example of connecting the theoretical assessment concepts in curriculum and practices in language assessment & evaluation courses to develop students' professional knowledge (Schulman, 1987).

In Turkey, the ELTE course became compulsory in teacher education programs in 1998 (Hatipoğlu, 2017). In this context, research undertaken by Şahin (2019) compared the syllabi & learning objectives of ELTE courses in Turkey, while studies conducted by Büyükahıska (2020); Hatipoğlu (2010, 2015) examined preservice teachers' needs and evaluations of ELTE courses while Öz & Atay (2017) focused on the relationship between classroom assessment and assessment knowledge.

To begin with, Şahin (2019) investigated the structure of syllabi in ELTE courses in state universities around Turkey. She has found out that the mainstream assessment and testing coursebooks and fundamental concepts determine the scope of ELTE courses in Turkey. Secondly, it has been emphasized that courses generally emphasize the summative role of assessment by engaging students in theoretical aspects of assessment, e.g., language skills & test types and principles of tests.

According to Davies (2013), curricula tend to be structured around constructing a "knowledge base (relevant background), principles (proper use of tests), and skills (techniques to develop tests)" in assessment courses. However, Şahin (2019) has inferred that the preservice teachers lack skills in the following areas:

- Adaption of materials to the specific teaching context
- Format & design problems in the structure of exams
- Lack of practice in constructing valid test items in different language skills
- Use of alternative classroom assessment tools

In conclusion, overdependency on traditional assessment approaches in preparing the course content results in a conventional form of testing in the narrow sense of "mere testing or grading" (Köksal, 2014). In fact, students should be well educated and 'assessment literate' to conduct assessment-related tasks rather than relying on ready-made, standard tests prepared by international publishers or testing units at schools (Haznedar, 2012; Şahin, 2019).

To identify preservice teachers' beliefs, needs, and expectations in designing the syllabi in ELTE courses in Turkey, Hatipoğlu (2010) worked with 124 senior ELT students at a state university in Turkey. She mentioned that preservice teachers demand the inclusion of practical topics in their L2 language testing and assessment course by considering factors such as assessment culture, local context, and students' career plans. Similarly, her needs analysis survey study in 2015 also reported similar results. It stated that a single course on L2 language assessment and evaluation is not enough

to prepare prospective English teachers to undertake the role of 'assessor' and confront many challenges in their future careers.

Similarly, a more recent study by Büyükahıska (2020) interviewed 39 senior ELT students at a state university in Turkey to uncover their perspectives on the course content. It has been found that, after taking a single ELTE course, preservice teachers do not feel well-prepared to undertake assessment-related responsibilities in their future profession and are ready to use sound assessment tools in their classrooms.

Öz and Atay (2017) investigated 12 EFL teachers' perspectives on the relationship between classroom practice and assessment knowledge. The findings demonstrated that teachers have excellent expertise in theoretical aspects of assessment, yet they fail to reflect their assessment knowledge in their classroom practices. They are far from utilizing formative assessment methods such as self-peer feedback, portfolio, presentations, discussion, interviews, etc., by which both students and teachers involve in the design of learning goals and implementation of assessment together (Brookhart, 2010).

Lastly, Mede and Atay's (2017) study examined the language assessment literacy and 'perceived needs' of 350 EFL teachers at state universities in Turkey through a language testing & assessment (LAT) questionnaire adapted from Vogt and Tsagari (2014) and semi-structured interviews. According to the results of the study, even though teachers did not consider themselves incompetent in second language assessment, the results of the questionnaire revealed that they are in high need of further training in LAT, which was provided by testing units at preparatory schools only in the writing/speaking sections of the exams. Furthermore, teachers reported that their preservice ELTE course and in-service training were insufficient. They also underlined that their reliance on “ready-made tests” and summative assessment is an 'institution-mandate.'

Furthermore, one of the teachers added that since they only have a 5% classroom performance grade on the student's final achievement grade, teachers' use of formative assessment tools in class is not willingly accepted by the students. Because the students

only aim at getting the minimum score to pass the English proficiency exam and start their undergraduate studies.

"[...] Students are not willing to do presentations when they hear that they won't get a grade. Sometimes I want to do something new, a communicative task or a discussion of a cultural issue, but get a reaction right away (EFL teacher, interview data, 11th April 2016 taken from Mede & Atay, 2017).

As Hatipoğlu (2015) stated, in exam-oriented cultures such as Turkey, teachers adapt their teaching practices according to the norms governed by stakeholders such as school administration, test experts, policymakers, etc. (Ricci et al., 2018). Therefore, they mostly stick to the traditional assessment methods as their colleagues while conducting their assessment practices (Hatipoğlu, 2015).

In this context, even though the curriculum reform in the second language assessment & testing course in 2008 by HEC and MONE emphasized the use of formative assessment and communicative language testing tools, educating teachers in line with this objective was not fully achieved (Hatipoğlu, 2017). The findings demonstrated that although the scope of curricula in ELTE courses widely covers all fundamental topics in assessment, there is still a gap between theory and practice. Lack of practice in the selection of materials, creating and adopting tests, item construction, and item analysis which are gathered under the term "hands-on experience" by Brown and Bailey (2008), are among the topics inadequately addressed by the course instructors as well as school administrations due to different local, instructional, and nation-wide constraints.

In this sense, the studies focused on conveying the needs of preservice EFL teachers remind us that most teacher education programs in Turkey still fall behind in designing the ELTE course in a way that meets the needs of preservice teachers in second language assessment and testing, which is a "notoriously difficult domain" (O'Loughlin, 2006) in teacher education programs (Hatipoğlu, 2010).

To make the necessary transition from theoretical concepts of assessment into practice, preservice teachers must "take the role of a tester" and develop authentic assessment

materials appropriate to their student's level and teaching context for their future profession. Therefore, the ELTE courses' content must be reformed to direct preservice teachers to practice language test construction and test-item writing because they will perform these kinds of assessment-related tasks very soon in their profession (O'Loughlin, 2006).

From this standpoint, the study carried out by Can (2020) serves as a model of how interaction in test item reviewing practices provides teachers with immediate and constructive feedback and thus increases the validity of test items, thereby language tests. In her study, Can (2020) investigated the item reviewing interactions of English language teachers at the testing unit of an English preparatory school in Turkey. Through the analytic perspective of conversation analysis, she examined how the structural organization of item reviewing practices evolved into problematization and eventually suggestion practices through teachers' collaborative work. During the problematization and suggestion phases of the item-review interactions, Can (2020) found that in-service teachers orient to the violations in the test items from different perspectives such as teaching perspective, students' perspective, and, interestingly, testing principles and rules. It has been discovered that teachers display orientations to the testing principles in constructing the test items. Therefore, they orient to the testing principles to provide an account for their problematizations as an interactional resource. In her study, Can (2020) discovered that professionals base their claims on professional grounds. In this respect, the study also demonstrates how teachers achieve mutual understanding and make joint decisions while improving the test item's validity and quality, which can critically affect students' lives in terms of making life-altering decisions in the educational context (Doughlas, 2010).

Another study conducted with preservice teachers in an online task-interface designed and implemented by Balaman (2015a, 2016) has shown that, preservice teachers mutually orient to the task rules set by the instructor throughout their interactions in an online task environment. It has been discovered that reservice teachers also negotiate the construction of new rules regarding the use of the second language for their task accomplishments. As first coined by (Amir & Musk, 2013) as “rule-policing” Balaman uncovered that, preservice teachers oriented to these rules to prevent communication breakdowns resulting from a breach of the 'second language

policy'. Therefore, it is seen in the study that preservice teachers added the rule policing to their repertoire as an interactional resource during their online task interactions. In another study conducted by Duran (2017), it has been discovered EFL students employ rule-policing 'as a situated practice' to deal with the knowledge gaps arising in their student-initiated practices, which leads to the utilization of both their first and second language at the same time. In this sense, Duran (2017) has demonstrated that the 'classroom cohort' and interactional needs of students lead them to adopt a language rule policy that promotes successful communication in a second language. Although these two studies take place in the EFL classroom context, they have put forward the conditions in which rule-policing has been utilized as an interactional resource by the students in its immediate learning context.

2.3. Peer Feedback as Formative Assessment

The rationale behind testing and assessment is its impact on educational management, which encompasses the planning, organization, and administration of pedagogical decisions (Ho, 2010) that shape students' future career choices.

According to Becker (1995), assessment as a fundamental component of teaching and learning can take two forms: summative and formative. Summative assessment, the conventional and dominant method, focuses on grading and placing students at all levels for the school year. In contrast, formative assessment concerns continuing learning that monitors students by showing their weaknesses and strengths for further action. In this regard, summative assessment is interested in evaluating students (Shepard, 2005) by grading, and it is still a dominant assessment method practiced by a large body of stakeholders around the world.

However, Rowntree (1987) draws attention to the educational purposes of evaluation, formative assessment rather than utilizing assessment solely as a means of grading in the traditional sense. He specifies the main aims of the assessment as selecting appropriate materials, controlling the quality of the education, and motivating students in their studies, and helping teachers in their lesson plans. But most importantly, closing the gap between students' current and intended performance (Sadler, 1989) so

they can manage their learning process effectively. In this sense, feedback is an integral part of the learning process to support students in their studies by providing necessary insights for their performance or work.

However, a considerable amount of literature (Bailey & Garner, 2010; Boud & Malloy, 2013) focuses on the problematic nature of feedback delivery and its dissatisfaction. According to a previous study by Nicol (2010), students may not fully benefit from the teacher's feedback as the absolute authority in class. Several studies (Crook et al., 2017; McConlugue, 2020) suggest that students' dissatisfaction arises from either lack of constructive feedback or receiving feedback in a technical sense, which is untimely, often vague, and non-specific. On the other hand, problems such as the management of overcrowded classrooms and students' failure to utilize feedback are also noted as difficulties faced by the teachers during the feedback process. For these reasons, it can be concluded that teachers and learners need a contemporary assessment approach and peer feedback to create a shared atmosphere in which learners become involved and responsible for their learning besides teachers. In this sense, peer feedback is the latest formative assessment model.

Therefore, Falchikov (2005) defines peer feedback as learners' judgment of each other's performance according to a standard criterion by directing questions, solving problems, and bringing solutions to the work under examination. By taking the role of an assessor from the teacher, learners actively become agents of their learning that foster lifelong skills such as self-autonomy, critical thinking, and self-efficiency, which are necessary qualifications for their professional future careers (Boud & Falchikov, 2006). In this regard, peer feedback becomes a popular and student-centered (Biggs & Tang, 2011) approach that is against the traditional assessment approach called 'assessment as measurement' (Serefini, 2001), which can degrade learners into the position of passive learners (Mobre & Teather, 2012). Suffice it to say, the advantages of peer feedback cannot be narrowed down to aspects related to the state of learners. What is more has been listed (Spiller, 2012) as the benefits of peer feedback as follows:

- encouraging cooperation among learners

- identifying gaps in the learning process
- enhancing analytical skills by providing and receiving feedback
- decreasing the power imbalance between students and teachers
- assigning learners as active agents of their learning
- managing time effectively and sharing the workload with teachers

In this sense, the last three decades have seen peer feedback as an alternative formative assessment method in higher education contexts in different learning settings.

Looking into these recent studies, one can say that a bulk of research has been conducted on the use of peer feedback as a popular assessment tool at the tertiary level from multiple perspectives, such as efficiency (Hattie & Timperley, 2007) on learning (McConglugue, 2020), organization, and implementation (Evans, 2013; Nilson, 2003; Pekrun, 2005), as well as benefits (Falshikov & Boud, 2007; Topping, 2017) and perceptions of learners (Azarnoosh & Huisman, 2018; Nicol et al., 2013; Sato, 2013) and teachers (Metin & Özmen, 2010; Sierra, 2015). In this regard, peer feedback has been widely applied in different learning contexts starting from EFL/ESL settings in writing classes to the L1 context in social and science classes all around the world. (Mendonça & Johnson, 1994) Furthermore, the research in pre-service teacher education has also addressed the implementation of peer feedback in the preparation of lesson plans and class observations in the following points: effect of peer feedback on teaching performance & development of teaching skills.

To begin with, studies conducted by Falchikov (1995, 2005) and Topping (1998) form the basis of studies focusing on self and peer feedback in the higher education context, which indicates a transition shift to the paradigm of social constructivism. Firstly, Falchikov (1995) conducted a mixed-method study with 13 undergraduate students in a psychology class in which students were assigned to do an oral presentation task. At the end of his research, Falchikov (1995) found that, unlike the traditional assessment methods, students provided each other with more informative and specific feedback.

Besides, the study aimed to test whether the main objectives, such as increasing self-autonomy and analytical skills, would be met at the end of the process. It has been reported that the goals of the study have been achieved. However, further planning and training students before implementing peer feedback are essential if teachers frequently want to use alternative assessment methods in higher education.

Moreover, Topping (1998) has brought evidence of the benefits of peer feedback in another L2 context. According to the results, students who gave peer feedback on their peers' L2 writing assignments found the feedback they received more understandable in terms of language & content, which yielded better performance and high grades.

The previous studies showing the effective use of peer feedback in the higher education context highly focused on how to implement formative assessment methods in higher education.

In this respect, scholars heavily focused on implementing peer feedback in writing classes in the EFL context (Lee, 2015; Liu & Sadler, 2003; Nelson & Schunn, 2008; Yu & Lee, 2016) and ESL context (Guardodo & Shi, 2007; Rollinson, 2005). The studies were mainly carried out in quantitative design and evaluated the use of peer feedback from different perspectives. Although the literature has multiple findings in implementing feedback, they found similar results regarding teachers' and students' positive attitudes. Furthermore, when the procedure is well-planned, the studies concluded that utilization of peer feedback in higher education increased learners' awareness of their weaknesses in L2 writing and pushed them to make necessary revisions for their final project. In this respect, recent but a few qualitative studies conducted in Turkey had examined perceptions of students towards their implementation of peer feedback in their L2 writing class at English preparatory courses and at state high schools in a similar vein.

According to Duruk (2016), English preparatory school students' perceptions of peer feedback changed more positively when they were also involved in planning the peer feedback in their classes. Kayacan and Razı (2017) and similarly Khalil (2018) also reported high school students showed better results in their writing assignments after

they received self and peer feedback revisions which were increased in quantity and rich in quality.

On the other hand, Demirel carried out a more comprehensive study in 2016. She researched 57 Turkish EFL students over 15 weeks in an L2 writing course at a state university in Turkey. In her research, she compared 1197 essay revisions in terms of organization, content, and form provided by peers and teachers. As a result, she found that modifications were high in number if given by the teacher and nevertheless more effective in content.

Lastly, the questionnaire on learners' perceptions of the different types of feedback revealed that students more positively welcome peer feedback in tandem with teacher feedback.

Another context researchers explored was peer feedback at secondary schools in science and social science courses at an international level. In the last decade, research has been conducted in mathematics (Calkins et al., 2020; Ross, 1995; Taşpınar & Halat, 2009), physics (Bulunuz & Bulunuz, 2013, Hansen & Andree, 2019; Tasker & Herrenkohl, 2016), and geography (Jenkins et al., 1994; Metin & Özmen, 2010; Morawski & Budke, 2019) courses at high school level all pointed to the use of peer feedback in teaching students' necessary theoretical concepts, conducting experiments, solving problems and increasing their engagement and unmasking their potential throughout their studies.

2.3.1 Peer Feedback and Preservice Teacher Education

However, there has been little discussion about the implementation of peer feedback by preservice teachers, who constitutes the most significant population among the stakeholders in the higher education assessment context. When considering the primary duties of a teacher, it has been widely accepted that teachers spend a significant amount of their time working on classroom “assessment-related work” (emphasis in original) to check the quality of education and achieve favorable learning outcomes (Crooks, 1988; Stiggins, 1999). In this sense, pre/in-service teachers utilize different assessment techniques to evaluate, monitor and support their learners

throughout the learning process. In this regard, preservice teachers' implementation of formative assessment approaches and their perceptions must be studied extensively. Therefore, in the last three decades, the interest in closing this gap in the literature deepened considerably. The studies investigating the learners' stances on peer feedback and its impact on the development of professional knowledge in preservice teacher has increased in number.

To start with the effective implementation of peer feedback, Ratminingsih (2017), Nguyen (2016), and Yüksel and Başaran (2020) found almost similar results in their research to understand the impact of peer feedback on the attainment of professional knowledge and fostering reflection. In this regard, Yüksek and Başaran (2020) examined 100 preservice EFL students in their final year teaching practicum to understand the effect of peer assessment and peer feedback as a reflective practice with a mixed method designed for the research study. As a result, it has been found that peer feedback enhances preservice teachers' professional knowledge and increases their critical thinking skills during peer feedback sessions. Because it has been found that while preservice teachers reflect on each other's performance, they have directed themselves questions, analyzed vital points, and, most importantly, held meaningful discussions. Through these procedures, preservice teachers negotiate meaning which helps them co-construct knowledge and become more active in their learning (Nicol, 2010).

Nonetheless, research also revealed that during peer assessment, preservice teachers showed subjectivity toward their peers with whom they are close friends when grading their performance which brings evidence of the involvement of personal factors in the process (Nilson, 2013). In this regard, the study also adds to the previous literature (Frunza, 2014; Torrance, 1993) centered on subjectivity and partiality as well as personal factors that can undermine the validity and reliability of formative assessment methods. Likewise, Azarnoosh (2013) compared to peer and teacher grades of 38 English literature students at a state university in İran in her mixed-method study to examine the effect of friendship on the assignment of grades. She concluded that although there is no significant relationship between the students' and teachers' grades, learners' attitudes towards peer feedback are highly positive. Still, learners must be

well trained before they benefit from peer feedback to eliminate any interference from environmental or personal factors that can undermine the validity of the assessment process.

Studies from different countries have also been conducted to fully grasp the personal characteristics and other factors that can affect the implementation of peer feedback in the same context. For instance, the study carried out by (Canabate et al., 2019) examined the views of 200 preservice teachers on the deployment of peer feedback during their micro-teaching practicum in the EFL context in Spain. The result of the mixed-method study illustrated that the motivation of preservice teachers from different majors had shown high correlations with their perceptions of autonomy and active involvement during their collaborative work in peer feedback tasks.

2.3.2. Perceptions towards the Implementation of Peer Feedback in Higher Education Context

Considerable attention has also been given to the key factors shaping the students' behavior. To exemplify, these affective factors are learners' motives, goals, beliefs, and perceptions, which are all shaped during their higher education (Topping, 2017). In addition, students' proficiency level in the target language and experiences gained from previous training within the specific learning context, EFL, and most importantly specific assessment culture have also been studied to understand the underlying mechanisms surrounding the peer feedback in the EFL higher education context. To begin with, proficiency in English has been an issue that all researchers agree on when deciding on the involvement of low-level and high-level students in peer feedback. In this regard, the review of studies illustrated that proficiency in English plays a preeminent role in the active participation and involvement of learners at different proficiency levels (Hyland & Hyland, 2006; Nelson & Carson, 1996) in peer feedback interaction. While it presented a challenge for low-level students, it offered a two-way interaction for high-level students to construct meaning cooperatively. Alongside proficiency, other components such as beliefs, perceptions, and particular motives of students are influential in their engagement. (Huisman et al., 2019). In this regard, the study conducted by Yu and Lee (2014) revealed that the students with the same proficiency level in English differ in their level of participation. These factors are

contextual and personal motives shaped during their peer feedback interaction in the L2 writing course, which yielded positive developments at different levels. A similar study (McCarthy, 2017) compared the attitudes of international and local students towards a triadic feedback delivery model given by staff, peers, and tutors in a classroom context or an online platform for two subsequent semesters in an L2 writing class in a higher education context in Australia. The results showed that many students valued the opportunity to receive feedback from varying sources.

Nevertheless, feedback provided by the staff was found to be more popular and favored by the local students than the two other feedback models. The local students had distinct attitudes towards different feedback types. They viewed the staff as experts who were well experienced within the field. While the international students employed online peer feedback more efficiently because it offered them a more detailed and in-depth feedback rather than the one, they could receive in-class sessions.

Likewise, Vickerman (2009) conducted similar qualitative research to understand the standpoint of 90 undergraduate students in the UK who are involved in peer feedback sessions as an alternative assessment method for the first time during their studies. The findings illustrated that a significant majority of students show a positive orientation toward the use of peer feedback as a contemporary approach. However, Vickerman noted that individual learner styles also significantly influence students' active involvement in feedback sessions. He reported that individual learners who are more self-autonomous might not utilize peer feedback as easily as others because they may find self-assessment a better approach that suits their characteristics.

However, after decades of work on peer assessment, it has remained a field scarcely explored in-depth (Kollar & Fischer, 2010). Because studies conducted so far have focused mainly on quantitative and qualitative research, which is based on questionnaires, and interviews of learners about their perceptions of peer feedback in the higher education context. However, studies describing the essence of student learning and the meaning negotiation and construction of knowledge in peer feedback implemented learning contexts have never received the attention they deserved.

In view of this, the study focuses on interactional practices of preservice teachers in a higher education context, second language testing, and evaluation course to display how preservice teachers develop authentic language tests, review the test items, and provide feedback to their peers through the microanalytic perspective of conversation analysis.

Existing research has been mostly restricted to addressing preservice and in-service teachers' needs and beliefs on language testing and assessment. At the same time, a bulk of research also focused on the assessment literacy measurement. However, far too little attention has been paid to understanding how preservice language teachers develop their assessment literacy and how they put this knowledge into practice in and through peer interaction when they are put into 'real-life situations in the classroom context.

As Davidson and Lynch (ibid) highlight, “test-crafting is not a linear process but a dynamic mechanism through which interactants in groups negotiate and resolve problems related to test items to obtain high-quality language tests that can serve the purpose of assessment” (p. 246). In this regard, to be able to understand how English language preservice teachers acquire necessary assessment skills through the enactment of social practices in peer feedback interaction and how this peer feedback dynamic facilitates learning opportunities for preservice teachers, analyzing micro-details of peer feedback interaction is vital. Therefore, this study aims to fulfill the gap concerning preservice teacher education and second language assessment in a peer feedback classroom interaction.

CHAPTER 3

METHODOLOGY

This section focuses on the introduction of Conversation Analysis (CA) as the analytical methodology of the current study, alongside its data collection and analysis procedures. In what follows, the validity and reliability issues surrounding the analytical method of CA will be addressed. Lastly, the research context and the participants will be presented.

3.1. Conversation Analysis

The current study adopts conversation analysis (CA) as the analytic method of inquiry. Conversation analysis is defined by Hutchby and Wooffitt (1988) as "the study of talk" (p. 7). Fundamentally, CA is interested in social actions accomplished by participants through the medium of talk and other bodily behavior (Sidnell, 2010). CA has its roots in Garfinkel's Ethnomethodology and Goffman's Sociology and comes into the picture as a "science of social action" (Drew, 2005) thanks to the collaborative work of Emmanuel Schegloff, Harvey Sacks, and Gail Jefferson in the mid-1960s in their pursuit of bringing a new analytical methodology to the investigation of social behavior (Stivers & Sidnell, 2013) that had been long analyzed under the dominant influence of cognitive and empirical methods in social sciences. The interface between CA and the disciplines that developed out of the studies led by the pioneer Sacks and his colleagues lies in the use of actual talk, talk-in-interaction, as the basic unit of analysis, which radically separates it from other scientific research methodologies that focus on empirical methods to study social behavior at that time (Ten Have, 2007).

CA's revolutionary perspective in the study of social behavior through the analysis of utterances and linguistic forms that were divergent from the viewpoint of dominant,

empirical, and laboratory-based experiments attracted the attention of linguistics and created a hot debate among linguists. Back in the day, the common understanding rested on language as being flawed and thus cannot contribute to the field of linguistics as stated by Chomsky (1965) confuted by CA. At the time when a unique yet analytical methodology developed thanks to the booming of technological tools, e.g., audio and tape recordings to demonstrate that it does not only aims to study conversation per se, but it seeks an answer to the question; “Why this, in this way, right now?' (emphasis in original) which places the focus on social action and its construction through talk-in-interaction (Seedhouse, 2004b).

Happen to talk about its background and emergence; a researcher must be familiar with key principles of CA. That is to say, (Seedhouse, 2005, p. 158) specifies the basic tenants of CA as follows:

i. There is an order at all points in the interaction

ii. Contributions to interaction are context-shaped and context-renewing

iii. No order of detail can be missed a priori as disorderly, accidental, or irrelevant

iv. The analysis is bottom-up and data-driven

Contradictory to the Chomskian view of mundane talk, interaction is systematized and structurally ordered yet not inconsistent or flawed. Secondly, interaction is context-bounded, which remarks that interaction is shaped in its immediate and dynamic environment and has the potential to reshape what follows up next (Sacks, Schegloff & Jefferson, 1974). Furthermore, as the third principle requires not missing any details and being able to analyze the organization of utterances systematically as CA dictates, the transcription (Jefferson, 1984, 2004) of the interaction must represent every detail of the actual talk as it is (Sidnell, 2010). Further to that, every detail of the interaction, which builds not only language but also body gestures, gaze as well as prosody, must not escape from the researcher’s notice since staying loyal the reality is a requirement while dealing with the video recordings capturing the microdetails of interaction (Liddicoat, 2007). The final principle puts forward the necessity of an emic perspective

which forms the basis of the conversation analytical approach. The emic perspective suggests researchers must not base their research on any prior presumption but their findings. Hence, descriptions of social behavior must originate within the naturalistic data set. Therefore, unmotivated looking must be the ultimate guide for researchers throughout the careful and repetitive examination of the data.

In short, the study of language as being the medium of communication and consequently making its way for interaction has the potential to get hold of the micro details of the conversation and human social conduct, unlike other methods which have failed to bring an account to the mechanisms surrounding the natural human behavior.

In that case, if one intends to make sense of the social dynamics of its environment, where to start and how to start? How does a researcher conduct research based on a conversation analytic approach and understand the interactional organization of conversation? First, one should get familiar with the concepts of adjacency pair, turn design & turn-taking, preference, and lastly, repair to fully grasp the analytical framework of CA (Schegloff, 1968, 2007).

To start with the basic unit of sequential organization, an *adjacency pair* (AP) is essentially a combination of two paired actions positioned successively (Schegloff, 2007) in which, upon the production of the *first pair part* (FPP), the *second pair part* (SPP) becomes relatively pertinent as what follows next. Let's take a request to exemplify. The request action initiated by the first speaker constitutes the FPP and requires a response as relevant in the next turn. Yet, first *preference organization* (Pomerantz, 1984) must be pointed out here.

The *preferred answer* to a request is expected to be the granting of what is demanded but naturally not the rejection as the *dispreferred answer*. At the same time, this exchange takes place in turns where the *turn-taking* mechanism serves as a fundamental basis in the organization of turns, turn-at-talk. Turns on the basic level consist of "a single word, a phrase, a clause, and ultimately a sentence" (Sidnell, 2012) and other features of prosody, bodily behavior such as gaze, laughter, etc. (p. 77). Hence, by dint of the components listed above, a turn is built from *turn-constructional*

units (TCU) which aims to accomplish a social action (Sacks, Scheglof & Jefferson, 1974). To instantiate our request, *Can you help me, please? I just can't lift the suitcase* consists of two TCUs hence by the type, it is a *multi-unit-turn*. Upon the composition of this TCU in the form of multi-unit turn speakers, project the upcoming turn-completion point where they signal that they will yield the floor in the turn-transitional relevance place (TRP). The interactants reveal TRP by employing many sources such as syntax, diverted gaze, rising intonation at the end of the utterance, please or outbreaths, etc. At the transition-relevance place (TRP), a change in the speakership becomes pertinent for the coparticipant to claim the next turn rightfully and timely.

As illustrated above, turn-taking and turn-design concepts represent the underlying mechanism of how participants' utterances are connected and employed to achieve mutual understanding collaboratively (Schegloff, 2007). However, there is always room for possible misunderstandings or hearing problems resulting in communication breakdowns which is a natural part of the mundane talk. To overcome these arising problems, interactants orient to repair practices that differ in design and turn irrelevant to their nature (Liddicoat, 2007). Firstly, if the ongoing speaker orients to a trouble source and brings a solution to fix the trouble source, it is *self-initiated self-repair* (Sidnell, 2012).

In contrast, the problem might be fixed by the recipient, and it happens to be a *self-initiated other-repair.* On the other hand, recipients might identify the trouble source and repair it themselves in *other initiated-other repairs.* Lastly, when the recipient pinpoints the trouble source, but the first speaker conducts repair, it is named as *other-initiated-other repair.* Still, in preference organization, self-initiated repairs are more preferred as a normative action other than other-initiated repair resolutions.

3.2. Transcription, Building a Collection, and Data Analysis

As Heritage (1995) states, “actual talk occurring in actual context” constitutes the primary source of data in CA and is defined by (Markee, 2000) as follows: “The analysis of talk-in-interaction minimally requires the use of audio or, preferably, video, recordings of participants' talk to capture the tremendous complexity of conversational

behavior" (p. 64). These recordings constitute the primary source of data used by conversation analysts. Starting with available technology, which is audio recordings in the mid-1960s, the researchers understand that video recordings are preferable for data collection because they capture the participants' gestures and gaze orientations alongside mere talk to describe the granular nature of the interaction fully.

With the same concern, the data collected for the current study employs video recordings from 3 different angles located in the classroom setting for six subsequent weeks to fully catch the verbal and nonverbal orientations of students, such as gaze, facial expressions, hesitations, and silences between turns, laughter, and pitch. Collecting the data with the video recordings enables the researchers to obtain the granularity of the interaction, which is required for the transcription of the data without ignoring any details. According to Wong and Waring (2010), the collection of the data by utilizing the present technology is the first principle of doing CA; the second principle is the transcription of the collected data, which is not a substitute for the recordings yet a necessary step to build an archive of the interaction before stepping into the analysis phase (Markee, 2000). As Seedhouse (2005,) puts it, "transcripts are inevitably incomplete, selective renderings of the primary data which invariably involve a trade-off between readability and comprehensiveness" (p. 251). Therefore 12 hours of the data is transcribed orthographically through the computer-assisted transcription software program, Transana, which was developed for managing qualitative data by Chris Fassnacht in 2001 to endow researchers with a more manageable platform instead of the tiresome manual method of transcription (Ten Have, 2007).

Later for the discovery of a possible phenomenon practiced by the students in their feedback interaction sessions, widely employed transcription convention developed by Gail Jefferson (1984, 2004) and Mondada's (2018) transcription system known for its integration of embodied actions such as gestures, face, and body movements have been applied. After the transcriptions offered a detailed representation of the data for analysis (Sert, 2015), I repetitively watched the data and read the transcripts within the emic perspective of CA for the analysis. Emic perspective refuses any pre-set assumption and solely depends on the data itself, and (Wong & Waring, 2010) describe

the rationale behind it as follows *Emic perspective* is a way of looking at language and social interaction from an "insider's perspective", i.e., stepping inside the shoes of participants to understand their talk and actions.". By closely examining the data multiple times and going through the detailed transcriptions within the boundaries of unmotivated looking, a possible phenomenon that is orientated normatively by interactants is discovered. It has been identified that; preservice teachers provide feedback to their peers by referring to the testing principles during their problematization sequences. It has been analytically diagnosed those preservice teachers give 'reference to the testing principles' while calling attention to the problems in the tests designed by their peers. The phenomenon is discovered with the use of 'should' structure when invoking the rules regarding the testing principles (See Figure 1) for instance, '`we should not eliminate the options by looking at the question`' however towards the final weeks of the summer course, they invoked the principles with a more fine-tuned use of language such as '`you can eliminate the first question therefore it's problematic`'. This can demonstrate that the principles might be added to their language assessment literacy.

Before starting my analysis, I have chosen an extract that represents the phenomenon under investigation to receive feedback from a group of experts in HUMAN, Hacettepe University Micro Analysis Network, in 2021. In the light of the feedback, I had been provided, I created my collection with 77 instances of the phenomena to form a general description, chose ten extracts, and assigned 3 extracts for each category to analyze within the framework of CA. I have assigned the term 'RTP' to the focal phenomenon in the study.

Table 1 Number of 'RTP' in Feedback Sessions

Sessions	Number of instances
Session 4: Grammar	33
Session 5: Vocabulary	8
Session 6: Reading	20
Session 7: Writing	17
Total	77

In the next chapter, the validity and reliability issues surrounding the CA methodology are addressed.

3.2.1. Reliability

Seedhouse (2004b) remarks that CA's employment of emic perspective to analyze naturally occurring data requires different techniques to address reliability and validity compared to the methodology of other social sciences. To begin with, Perakyla (1997) mentions the main concepts that can address reliability issues surrounding CA by paying attention to the quality and abundance of the visual recordings for the collection of the data and the granularity of the transcriptions. In this sense, the study is composed of 12 hours of video recordings transcribed through the Transana software program and elaborated with the transcript convention systems of convention systems of Jefferson (2004) and Mondada (2018).

Nevertheless, the question raised by Ten Have (1999 as cited in Seedhouse, 2004b) is whether the findings from naturally occurring data are replicable or not? At this stage, the analytical framework of CA becomes an integral part of the discussion. The presentation of the data in CA studies is relatively different from other research methodologies that only disclose their analysis when publishing but not the data itself. However, in the case of CA, the researcher already must present their data through detailed transcriptions that allow readers to go through the analysis and come up with arguments that can approve or deny the results of it (Seedhouse, 2004b). In this case, an extract from the collection of the study was brought to an online data session held by Hacettepe university Micro Analysis Network in March 2021. In the data session, the researcher initially briefly presents the background of the social context in which

the interaction occurs. Then, a community of experts and researchers gather to analyze an extract with its transcription and repeatedly play video recordings from 3 different angles.

Participants in the group start their close examination of the data through the emic perspective, which is secured by the researcher's withholding the phenomena s/he works on to avoid interfering with the participant's point of view. Researchers share their observations by actively participating in the discussion, which is summarized by Weiste and Stevanovic, (2017) as " taking a turn equals " making a CA contribution. By holding analytical discussions with the research group members described above, the researcher receives feedback and collaboratively increases the study's reliability and validity.

3.2.2. Validity

Brink (1993) states that validity and reliability determine the quality and credibility of the research. In this respect, a popular question directed to CA in the social sciences paradigm is how one can ensure validity while dealing with human behavior. First, we must address four types of validity: internal, external, ecological, and constructive validity. To begin with, internal validity deals with the trustworthiness and reliability of the findings. It raises the question of how one can justify the conclusions of a conversation analytic study. The methodology of CA brings an account to this question with the emic perspective; that is, researchers do not hold any preestablished views when making sense of the interactional practices of interactants but rather describe their orientations from the participants' perspective. In this sense, CA researchers can only base their analysis on the details provided by the interaction itself but no other factors such as context, identity, gender, etc., unless applied by the interactants themselves. Hence CA draws its strength from its emic perspective by integrating the perspective of both interactants and the analysts, maintaining its internal validity. Likewise, CA outperforms other social disciplines which draw on unnatural environments to collect their data by studying naturally occurring interactions whose results can be re-applied in different, daily situations to be generalized as a requirement of ecological validity. Similarly, external validity requires generalizing the findings to other contexts (Seedhouse, 2004b) apart from the context interaction.

In this regard, CA aims to put forward normative descriptions of interactants' practices in their unique context within the micro and macro-scale analysis. In this way, it justifies the rationality of the external validity in its analytical methodology. (Seedhouse, 2004b). Finally, construct validity in CA means human beings build their reality, id est, and constructs to make sense of their world due to their interaction. In other words, their constructions originate from their social actions, which come into being through talk-in-interaction.

To summarize, CA's emic perspective and micro-analysis of the details in interaction provide a unique advantage when it is compared to the methods employed by other research methods in social sciences to address validity and reliability issues. In this regard, Merriam (1995) mentions several steps qualitative research methods take to ensure validity and reliability. She talks about the widely accepted member check method, triangulation, peer examination, and engagement in the data for long hours during data collection and after the analysis stages.

Likewise (Davis, 1992) states that “prolonged engagement and persistent observation” are essential while conducting research in social sciences. In the analytical methodology of CA, one cannot fully put these methods into practice considering the necessity of CA-practitioners and the granularity of the data. However, in the present study, I have presented a part of the data in an online data session to benefit from the contributions of a group of researchers in the field. I also engaged myself in transcribing the 12 hours of data determined as highly sufficient for conversation analytic research according to the criteria set by Seedhouse (2004b) and required me to go over it multiple times to create a representation of the data. Moreover, I have enjoyed the benefit of my two colleagues working in the same field while receiving their feedback and crucial insights from my advisor. Therefore, the present study has utilized various methods to validate its findings by practicing techniques intended for qualitative research and conversation analytic approach.

3.3. Participants and Research Context

The participants in this study are 23 senior-year preservice teachers in the department of foreign language education at a state university in Ankara, Turkey. Seventeen of the preservice teachers in the present study are female, and 6 of them are male. Their native language is Turkish, and they are competent speakers of English. According to the CEFR, to be able to start their undergraduate studies and be exempt from the English preparatory program at their university, they must pass the English proficiency exam with a minimum passing grade of 70 (equal to TOEFL 86 and IELTS 7.0). During their studies, preservice teachers have not taken any course related to testing and evaluation according to the curricula of their educational program, id est, ELT. Therefore, the participants took the English language testing and evaluation (henceforth ELTE) course as an obligatory core course in the final year curriculum of the English language education program. The course is offered to the senior year preservice teachers two times a year at different time intervals. In the 2018-2019 academic year fall semester, the course took 14 subsequent weeks extensively, and in the summer school, it was taught as an intensive course that lasted six weeks.

As (Hatipoğlu, 2015, 2017) puts forward, testing, evaluation, and assessment course is integral to the ELT curriculum. The courses offered at English language education programs aim to equip preservice teachers with the necessary skills (The American Federation of Teachers, The National Council on Measurement in Education, and the National Education Association, 1990) listed below:

- Selecting and developing appropriate and good-quality assessment techniques for the specific teaching context and learners.

- Conducting the test and interpreting the results by paying enough attention to the validity and reliability standards.

- Taking advantage of the results to develop necessary materials and plans for the school and the specific areas students fall behind.

- Providing feedback to the students, parents, and school administration to take further action.

- Being aware of the outdated and ineffective assessment practices and not perpetuating common mistakes in the field during the test construction & adaptation process

In the light of these standards to develop language assessment competency and literacy of preservice teachers, this study focused on the 2018-2019 summer school period, which adopted a flipped classroom model (See Table 2).

Flipped classroom model is basically designed in a way where the students complete their course readings and homework at home and come to the classroom prepared for participating in learning activities. As Bishop and Verleger (2013) summarizes "lectures and homework can be completed outside of the class, while active learning activities take place within the classroom." To present the implementation of the peer feedback in the context of this study, one should examine the table provided above. As presented in Table 2, the preservice teachers and the professor who teaches the course and ELTE met two times every week on Mondays for 3 hours and on Tuesdays for 4 hours. In the first two and a half weeks, the course covered a broad range of topics on the introduction of testing and evaluation, different types of international & national tests, test construction and item-writing process, and reliability and validity measurements.

Table 2 ELTE Course Outline

Week	Days	Topics covered in lecture	Emergence of the principle	Principle being referred	Representative Extract
1.	Monday	Introduction to the course	Validity & reliability	placement of the instructions	4a,4b,5
	Tuesday	Kinds of tests	Placement of the questions and exam sections	-	-
2.	Monday	Validity & reliability	-	-	-
	Tuesday	Writing & evaluating multiple choice questions	Constructing multiple choice items	Odd-one-out, avoidance of ambiguity	1, 6a,7,9a,9b 2
3.	Monday	Holiday	-	-	-
	Tuesday	Testing grammar and vocabulary	Designing distractors in multiple choice items	Test what you teach, Difficulty level	8a, 3,8c
4.	Monday	Testing reading	Writing and evaluating multiple choice items	-	-
	Tuesday	Testing writing	Writing and evaluating multiple choice items	-	-
5.	Monday	Testing listening and speaking	Evaluation of productive skills	Identification of acceptable answers	8b
	Tuesday	Testing speaking	Evaluation of productive skills	Avoidance of ambiguity	6b

In the meantime, preservice teachers formed peer groups to write the first drafts of their exams. According to Bishop and Verleger (2012), one of the most critical steps that must be taken before implementing peer feedback is to engage students in the

preparation process. Therefore, the design of the groups, submission of the exams, and the general scheme for the feedback sessions are all decided in full collaboration with preservice teachers taking the ELTE course.

The exams were prepared for the students in grades 5,6,7, and 8 and consisted of 6 main sections (vocabulary, grammar, reading, writing, listening, and speaking). (See Table 2). As of the 3rd week, the course implemented the flipped classroom model in which preservice teachers completed the assigned readings at home, prepared their exams in groups, and provided written feedback for the exams of their peers at the university's online learning management system, called ODTUCLASS. Each group provided feedback for two peer groups, respectively, and groups corrected their exams according to the received feedback. However, unlike the extensive course given during the fall semester for 14 weeks, preservice teachers had a limited time to make the most of the feedback they received.

Table 3 Peer Feedback Groups

Group Names	**Group Members**
Group 1	HAS, REY, BAR, OZG
Group 2	ELA, PER, MER, FEY, SOR
Group 3	TUĞ, İLK, UZE, NEH, OYA
Group 4	HAL, BUS, ZEK, GUL,
Group 5	MUS, MEH, AYS, NUR, SIB

All the groups are expected to present their comments for the exams written by their peers in terms of content, format, practicality, and validity. Moreover, pre-service teachers had to validate their comments through assigned course readings which consisted of books

and articles based on foreign language testing, evaluation, and assessment by the course professor. The preservice teachers were expected to make necessary revisions on their exams following the feedback they took from their peers and upload the revised versions on ODTUCLASS before the end of week 6. By week 6, the students had a presentation session, and they introduced the final version of their exams to the

whole class and reflected on their test-construction process by practicing self-assessment.

Table 4 The Duration of Recordings

Sessions	Session Type	Duration
Session 4	Feedback on the grammar section	186 minutes
Session 5.	Feedback on the vocabulary section	128 minutes
Session 6.	Feedback on the reading section	125 minutes
Session 7.	Feedback on the writing sections	122 minutes
Total		12 hours

The study only involved recordings from the feedback and presentation sessions of the data. Even though the whole data set is transcribed, the subject of the data is 5 classes which make up 12 hours and 728 minutes of the summer course in total.

CHAPTER 4

ANALYSIS AND FINDINGS

This chapter presents the analysis of the phenomenon “reference to testing principles” (RTP) from 10 extracts selected from the collection consisting of 77 instances by adopting conversation analytic approach. The analysis demonstrates the instances of ‘referencing to testing principles’ in three different sequential environments in peer feedback sessions as follows; order 1) problematization of test items 2) resistance 3) suggestion.

4.1. Employment of RTP in Problematization Phase

In the first section, the use of RTP during the problematization phase of the peer feedback interaction is presented in five extracts. The first extract is taken from the fifth feedback session of the summer school, which is dedicated to the vocabulary sections of the exam. However, due to time reasons, group 5 could not receive any feedback in the grammar session held on Monday. Thereby, group 5 receives feedback on their grammar section at the beginning of session five before the class continues with the vocabulary exams. In the extract below, group 5, consisting of MUS, MEH, AYS, NUR, and SIB, receives feedback from group 3, TUĞ, İLK, USE, NEH, OYA, and group 4, HAL, BUS, ZEK, GUL. The test-writer group prepared their exams for grade seven and received feedback on one of their multiple- choice questions whose options are problematic in terms of its design; therefore, it violates a testing principle in the construction of distractors.

Extract 1. Group 5/Options

1 T: **oka:y (4.0) so any problems with *the question**
*--1---> line 3

1:HAL turns the page and looks at BUS next to him

2 **>your two year-old sister [i:s**

3 ILK: **[e::r**

4 T: **<younger(.) #elder(.) taller(.) #older>***
#--2--->
--->*

2:ILK raises her hand

5 HAL: **°unintelligible it is +testing vocabulary°**
+---3--->

3:teacher looks at İLK and nods her head

6 ILK: **oka:y so: (1.2) 1ϒone se:cond (1.2) er:m**
ϒ---4---> line 8

4:İLK looks at her computer

7 **↑in the first question option b↑ contains word el:der**

8 **is %problematic% because it is a noun compared to**

9 **other ohp- options which are comparative forms of(.)**

10 **[nouns (0.3)ϒyanlış>**
--->ϒ

11 T: **+hmm hmm+ [hmm hmm**
+T nods head+

12 ILK:→ **options &shouldn't differ from each other in forms**
&---5---> line 12

5:ILK raises her head and looks at teacher

13 **(0.2) we said&**
--->&

14 T: **AGREE:: ♣A:ND BESI:DES↑ it is a::=**

--6---> line 17

6:T looks at classroom

15 BUS: **=[Ψit is a vocabulary=**

ΨILK nods her headΨ

16 HAL: **[+it is a vocabulary+**

17 SS: **[Σvocabulary**

Σ---7---> line 18

7:opens her hands towards BUS and nods her head

18 T: **[oka:y]♣**

--->♣

19 BUS: **↑it is Σa vocabulary question [because**

Σ--->

20 T: **[+th-+**

+t nods head+

21 BUS: **(0.2)~all of~ the options are structurally same(.)**

22 **it could be <than on the stem> (.) ↓but (0.2)**

23 T: **≈this is ∝not (0.3) A grammar question(1.1)**

≈--8---> line 27

8:shakes head repetitively

24 **< ∝this is a: vocabulary question πoka:y↑(1.2)**

--->∝ π--9---> line 28

25 **so you have to (.) write a new**

9:AYS and SIB nod head

26 **question (1.2) for the comparative (0.9)**

27 **#because everything is ↑there >the**
#--10--->

10:SIB take notes

28 **only thing and students need to kno:w< i:s understand**

29 **that the brother is two years o:ld and then to choose**

30 **>what is related to age and that's is(.) but that's**

31 **vocabulary okay so that i:s >that question is cross(.)**

32 **≈cross it out and then >start writing a again πokay(.)**
≈--->
33 **now question two.**

Extract 1 begins with the teacher's turn-initial (okay) which signals her readiness to initiate the feedback session with a (4.1) seconds gap that probably displays her wait time for preservice teachers to show participation, as it is evident in her gaze sustained at the classroom during her wait time. After no one claims the turn, she signals her transition to the next question on the exam with the transition marker (so). She asks any problems related to question one to elicit answers from the preservice teachers. She starts reading aloud the focal question by making it available to everyone in the classroom, during which İlk from peer feedback group 3 produces an elongated hesitation marker (e:r). However, the teacher does not show orientation to her as she is busy reading aloud the question stem while sustaining her gaze at the exam paper written by group 5. While the teacher continues her reading the distractors aloud, this time, İlk raises her hand and bids for a turn (Sahlström, 2002). At the same time, Hal from peer feedback group 4 turns to her group member Bus, makes a comment in sotto voice on the focal question, and states that the questions test vocabulary, not grammar skills which violates the testing principle' 'the test measures what it aims to test as a part of face validity. However, the teacher does not show any orientation to his comment, probably because he is sitting at the back of the class and hence inaudible.

On the other hand, the teacher sustains her gaze at the classroom for a possible attempt for the next turn and notices İlk's bidding for a turn, upon which she allocates the turn by nodding. In the following line, İlk takes up her turn with the marker (okay), which

signals her readiness to answer (DeSouza, 2021) the question directed by the teacher in line 2 (any problems with the). However, as she is orienting at her computer probably to find her notes for question one, she requests time (one sec). After (1.2) seconds of a gap, she formulates the base sequence of her problematization with a rising intonation. She asserts that the focal question is problematic through negative assessment marker (problematic) by directing attention on option b (word elder) among the other distractors, which are different in forms. She continues her turn with the discourse marker (because), which signals an upcoming account-giving to her problematization in the prior turn. In line 7, she leans into her computer and states that option b is a noun while the other distractors are comparative forms the test aims to measure. Following a micro pause (0.3), she produces an explicit negative assessment (yanlış/wrong) at a slower pace and looks at the teacher. During her turn in progress, the teacher nods and acknowledges (hmm hmm) her at multiple points. In line 11, İlk explicitly references the principle that 'options should not differ from each in forms (Fulcher, 2010; Hatipoğlu, 2009). In her account-giving sequence, İlk invokes a testing principle that she thinks violated in the design of the distractors in a multiple-choice question test designed by group 5. According to Waring (2017) and Can (2020), preservice teachers base their accounts on "solid grounds" to display their understanding and expertise in the construction of test items. While providing an account of the problematization announced by herself, İlk also positions herself in (K+) position, which is evident in the item-writer groups' orientation to her claims by nodding heads and taking notes (Heritage, 2013; Sidnell, 2012) until the end of the discussion. After a brief gap (0.2) seconds, she further adds with the inclusive marker (we said) to emphasize that the comment is made as collaborative group work and she is speaking as a 'representative on behalf of an organization (Drew & Heritage, 1992). In the following turn, the teacher displays that İlk's comment is a preferred contribution with her articulation of an elongated agreement token (AGREE:)in coordination with head nodding. In the same turn, she employs a discourse marker (besides) which signals an additional problem within the question and immediately produces a designedly incomplete utterance DIU (Koshik, 2002) with (it is a::)to elicit the additional problem from the preservice teachers. In the following turn, her formulation of DIU immediately receives multiple completions from the preservice teachers starting with Bus and Hal and later in choral co-production (vocabulary)

from the preservice teachers beginning at the final utterance of Bus and Hal. It is noteworthy to state that the additional problem in the test item is also noticed by other peer feedback groups and that they have oriented their role and responsibility as' feedback-providers' by reviewing the item (Can, 2020; Heritage, 2013).

Besides, in the teacher's formulation of a DIU, she includes all preservice teachers in the classroom, which makes it a multi-party interaction (Schwab, 2011). After eliciting the additional problem in the test item, the teacher opens her hand towards both sides and produces an acknowledgment token (okay) which indicates her alignment and agreement with the preservice teachers. In line 18, Bus self-selects and repeats her claim that the test item is a vocabulary question and issues a discourse marker (because) to signal her upcoming account-giving for the problem. However, her yet-incomplete turn overlaps with the teacher's turn in line 19. After the teacher nods at her, she takes up his turn again in line 20 and continues her account-giving sequence with rising intonation to state that all the other options in the question are structurally the same, which is in line with İlk's comment in line 6. Bus also remarks that the question stem has (than) which is the comparative structure the test aims to measure.

Nonetheless, she cannot provide the rest of her account as her falling intonation signals the end of her turn when she aborts it with the contrastive marker (but) where the teacher takes up in the next turn. In line 22, the teacher shakes her head and restates that the item is not a grammar question but a vocabulary question which is also another violation of the principle that the test measures what it intends to test. Between lines 22-28, the teacher employs an extended telling and issues an understanding check with the token (okay↑) . Even though she does not receive any verbal uptake from the item-writer group, SIB and Ays show their nonverbal alignment by nodding their heads multiple times during the teacher's formulation of her suggestion. After a (1.2) seconds of silence, she proceeds her turn and suggests writing a new question instead of the problematic one. In the following lines, she elaborates on the problem first stated by İlk in line 3 that all the distractors in the question are structurally the same, and the only option related to the age and comparative structure is given in only one option (option b, elder) which students can easily eliminate. In her elaboration of the problem announced in the design of the question, the teacher also enacts non-present

actors, students with the third person plural (they) by assuming them as the potential test-takers to formulate her problematization from the perspective of hypothetical students to describe a possible behavior they can show while taking the test (Leyland, 2021; Yöney, 2021). Besides, she further advises with her embodied gesture (Mondada, 2007) to cross the question out, which is formulated as a directive to the item-writer group who performs a claim of understanding through the acknowledgment token (hmm hmm) in coordination with head nodding after teachers' understanding check (okay↑).

Extract 1 displays preservice teachers' orientations to different testing principles, which are violated in the design of distractors constructed by group 5 to measure the comparative structure in the grammar section of their achievement exam. Concerning the 1) design of the options in different forms and 2) testing what it is intended, preservice teachers do not only announce the problems in the test items but also provide accounts of their claims by explicitly referring to the violated testing principles as an interactional resource. Therefore, from this moment on, preservice teachers' employment of testing principles as an interactional source will be referred to as 'Reference to testing principles' (RTP), which will be better described in terms of its functions performed in different stages of the peer feedback session. To exemplify, the first function of RTPs in the problematization phase of the test items is to bring an account to support the claims of peer feedback groups as presented above. The use of RTP in extract one shows that providing accounts before the initiation of the problematizations leads to the recognition of the problems and therefore establishes mutual consensus on the matter.

Furthermore, the fact that the teacher also shows alignment to preservice teachers' arguments by elaborating on the identified problems indicates that the problems are also noticed by the teacher and lastly recognized by the item-writer groups who display nonverbal orientation to the comments by taking notes and nodding until the end of the discussion. It is also important to note that, by bringing evidence to the announced problems through invoking testing principles, preservice teachers position themselves in (K+) epistemic status (Heritage & Raymond, 2005), as is evident in item-writer

group's taking notes which signal their realization of the problem due to the comments provided by the peer feedback groups.

Extract 2 is taken from the last feedback session of the summer school dedicated to the writing sections of the exams prepared by preservice teachers. In this extract, group 3, consisting of TUĞ, İLK, USE, NEH, and OYA, receives feedback on their design of the question, which requires students to form sentences using prepositions according to picture prompts about white goods in the kitchen. The exam is prepared for 6th graders and intends to test prepositions and numbers. At the beginning of the session, the group received feedback regarding the face validity problems in the test design and exam specifications from peer feedback group 4 HAL, BUS, ZEK, GUL, and group 5 MUS, MEH, AYS, NUR, and SIB.

Extract 2. Week 7- Group 3/ instructions (05.45/ 7.31)

1 HAL: **=°hocam bir de dokuzuncu soru:da(.)**
miss there is also in question nine

2 **instructionlarda noun clause var**
noun clause included in the instructions

3 **°unintelligible=°**

4 T: **=ohhkay say it (3.0) say it >out loud<**

5 HAL: **→ha şey e:r in the tenth question**
oh I mean

6 **(1.2) er: there is noun clause structure(.)**

7 **in the instruction (0.4) but there is**

8 **+no: (.) noun clause (.) +in the: (0.2) curriculum**
+------t nods head------+

9 T: **hmm hmm (0.4)**

10 HAL:→ **+instructions should be::(0.3) understandable °for**
+ ----------t nods head----------+

11 **students° (0.5)**

12 T: **hmm hmm ↑CORRECT (3.0)**

13 HAL: **%°ben° açıkçası şeye takıldım**

I to be honest something stuck in my head

%--1--->line 20 #--2--->

1: T looks at halil

14 T: **+hmm hmm+**
+nods head+

15 HAL: **look at the prices >ask and answer diyor ya<**

you now it says look at the prices and answer

16 **yani bu: yenisinde de(.)aslında aynı şekilde**

I mean in the new version as well actually the same

17 **sade:ce <write how much they cost in words> demiş mesela**

its only written how much they cost in words for example

18 T: **+hmm hmm+**
+T nods +

19 HAL: **şimdi burda be:nim(0.2) neyi yazacağımı**

now here my what am I going to write

20 **+ben [bulamadım yani+ (.) %şey ikinci**
I couldn't find you know I mean the second one

+---T nods head---+ %--->

21 BUS: **[°şey altına cevaplar lazım mesela**

you know answers are necessary below for example

22 **°write how much they cost [yenisin:DE**

in the new version as well write how much they cost

23 NEH: **[unintelligible]**

24 BUS: **[yenisinde bu**

25 HAL: **[YENISINDE DE şimdi mesela >refrigarator**
in the new version as well now for example

26 **costs two thousand turkish liras mı yazcam< (0.2)**
am I going to write

27 **yoksa: hepsini total mi [<yazıcam yoksa**
or am I going to write all of it in total

28 **hepsını aynı anda-°**
or all of them at the same time

29 BUS: **[it costs in total diYİP olarak da yazabilirdik bir cümle**

we could have written just a sentence

30 T: **şimdi şöyle mi yazıcam REFRIGATOR I:S(0.2)**
now am ı going going to write as in

31 **and then in words two thousand lira mı yazıcam**

32 **yoksa saDECE REFRIGARAT::OR iki nokTA::two thousand lira**
or am ı only going to write

33 **(0.2) [mı yazıcam]**

34 NEH: **[°unintelligible°]**

35 T: **&ne yapSIN (0.4)% hadi sınıfça nazlıya birazcık**
Come on what can she do as a class just a little

&HAL raises his hand--->line 35

36 **yardımcı olalım**
let's help her

At the beginning of the session, Hal self-selects and address the teacher (hocam/miss) through the discourse marker (bir de/also), which signals his upcoming problematization to pinpoint another problem in the writing section of group 3. Hal states that noun clause structure is included in the instructions, which is a grammar structure taught at high proficiency levels; therefore, it is not suitable for a test designed for 6th graders. In line 4, the teacher immediately acknowledges (ohhkay) while instructing him to say it louder as he sits behind the classroom and the last utterance of his turn is inaudible. In the next turn, Hal produces a change of state token (oh) which marks his realization (Heritage, 1984). He starts reformulating his initial claim with the preface (şey/well) by focalizing the tenth question, which includes noun clause structure. He further signals his upcoming negative evaluation (Can, 2020) with the contrastive marker (but) and adds that the noun clause structure is not included in the curriculum, which indicates the violation of the testing principle, "vocabulary, grammatical structure, or the statement is not studied in class, and hence students may not be able to understand it". Without a time-lapse, the teacher shows her alignment (hmm hmm) in coordination with an embodied head nodding. Hal sustains his gaze at item-writer Neh and proceeds his turn by referring to the violated testing principle in the construction of instructions by explicitly stating that "options should be understandable for students." By orienting to the principle, Hal provides a post account for his problematization in the previous turn regarding the writing of

instructions. According to Waring (2017), by invoking larger principles, teachers "convey or cement the understanding as a matter of fact impervious to questioning." Therefore, with an account-giving, Hal justifies his claim based on testing larger principles which is a set of rules guiding the testing and evaluation pedagogical discipline. Furthermore, it is evident in the formulation of his multi-unit turn that he positions himself in (K+) status (Heritage, 2013), as it is evident in the next turns of the interaction that item-writer groups respond and take notes which indicates their noticing of the problem because of HAL's comment in the previous lines which puts them in (K-) epistemic status (Heritage, 2013).

In the next turn, the teacher displays her alignment and agreement through the confirmation token `(correct)` with a rising intonation. However, Neh, the writer of the test item, does not orient to the announcement made by Hal. She only shows bodily orientation by sustaining her gaze at him during his problematization act. Following a micro-pause (0.5), Hal contacts mutual gaze with the teacher and takes up his turn again in line 13 with the turn-initial discourse marker `(açıkçası/as a matter of fact)` to elaborate on his problematization regarding the instructions as a post-expansion to his initial claim in line 3. By formulating the problem from the perspective of his experience with the item `(ben açıkçası şeye takıldım)`, he states that he had difficulties with understanding the instruction and further reads aloud the instruction written in the question stem to elaborate on how it is designed and how it creates further problems. In line 18, the teacher acknowledges `(hmm hmm)` while at the same time nodding her head. In line 19, Hal produces a deictic referential `(burda/here)` while pointing at the exam paper (Mondada, 2007) in front of him and expresses his difficulty in understanding what to write as an answer to the question. As a complementary to his initial claim concerning the instructions being not understandable, he displays a potential test-taker behavior by reading aloud the question stem and showing his difficulty with understanding the instruction `(ben ne yazacağımı bulamadım)` by referring to his test-taking process as an end-user of the test item.

In line 21, BUS, from the feedback provider group 4, shows her alignment and agreement with Hal's comment by expressing that there must be example answers

under the question stem (altına cevaplar lazım mesela), which shows her treatment of the lack of answers as a problem that fails to guide test-takers while answering the question.

This is met with a response by NEH in line 23, however, because of the overlap with Hal and Bus, her response is unfortunately unintelligible. However, in the next turn, both Hal and Bus respond to her by stating that they are referring to the last version of the test; therefore, it can be inferred from the next-turn-proof procedure that in her unintelligible turn, Neh mentioned something related to the old and new version of the exam. At the beginning of the session, the teacher reminded group 3 about the rule, which is to create the last version of the exam copies after the groups receive feedback. Because Neh brought two copies of the exam to the session and hence she received feedback on both versions of her exam throughout the session. In line 25, Hal extends on his claim and issues an alternate wh- interrogative to Neh. In his comment, he asks Neh whether to write the total prices of the items under the pictures or to write the price of each item individually as an answer. In his turn-in progress, Hal displays the problem with the instruction in the question stem by providing two possible answers, which indicates that the design of the instruction allows test-takers to come up with more than one possible correct answer, which violates another testing principle 'only one correct answer'.

Just before his completion, Bus overlaps with his turn in-progress in line 29. In what follows, Bus formulates another possible answer (it costs in total diyip de yazabilirdik) and displays her agreement with her group member HAL by challenging Neh. In the following lines, Neh does not provide any response to the challenges raised by feedback-provider groups; therefore, the teacher takes up in line 30, shows her affiliation and agreement with the comments of the feedback-provider group 4 by employing a full modified repeat of HAL's comment in line 26. During her turn, she restates the problem about multiple possible answers to the item because of the unstructured, general, unguided instructions (Heaton, 1990) given in the question stem that leads to confusion about the expected answer. It is noteworthy that, while providing possible answers to the item, preservice teachers Bus, Hal, and teacher frame their claims by taking themselves as potential end-users of the test items, that is, test-

takers. In the following line, item-writer Neh provides a response to teacher's inquiry about how to answer the question, yet her response delivered in a sotto voice, unfortunately is unintelligible. In line 35, to bring a possible solution to tailor the problematic instruction for sixth graders, the teacher addresses the whole classroom which makes it a multi-party interaction (Schwab, 2011) and thereby creates an advice-giving environment that eases the epistemic asymmetry between the teacher and the preservice teachers by inviting them to show participation `(ne yapsın Hadi yardımcı olalım/what does she do, let's help her).`

It is displayed in Extract 2 that preservice teachers orient to their role and responsibility as peer feedback groups by pinpointing potential problems in the design of the test items that can affect the validity and reliability of the language exams. While identifying the potential problems, it is seen that preservice teachers also refer to the violated testing principles causing further problems in the construction of the test item. During their reference to testing principles, Hal also brings evidence to the violated rule by positioning himself as a possible test-taker who might provide more than one possible answer to the question because of the unstructured, unguided, and vague instruction provided in the stem. His reference to the violated testing principle in coordination with his depiction of a possible test-taker behavior ensures mutual understanding about the nature and severity of the problem in the item, which received full agreement and alignment from his group members and the teacher. It is after Hal' problematization of the test item that the teacher and the preservice teachers in the item-writer group notice the problematic issue in the writing question. His problematization and account-giving to the problem through the testing principles leads to the realization of the problem and which puts item-writer groups (K-) position (Heritage & Raymond, 2005).

Extract 3 is taken from week 5, dedicated to the vocabulary section of the exams. In the extract below, Group 5, consisting of MUS, MEH, AYS, NUR, and SIB, receives feedback on their vocabulary section based on a definition and adjective matching exercise from group 4, HAL, BUS, ZEK, GUL, and group 3 TUĞ, İLK, USE, NEH, OYA.

Extract 3. Week 5 Vocabulary Session 'definitions'

1 T: **A:::neyse hadi şimdi başlayalım baştan tamam mı**
anyway let's now start from the beginning okay

2 **e:r kim demişti ϕvar mı sizin commentlerinizi**
who said do you have
ϕ--1--->

1: Hal raises his hand--->line 4

3 **dinliyim ondan sonra benim söyliyeceklerim çok**
I listen to your comments then I have lots to say

4 TUĞ: **°oh°**

5 T: +**SÖYLE ϕhalil+**
ϕ--->
+nods head+

6 HAL:→ **şim:di er ↑you told us (.) <that er**

7 **the definitions should be an-an minus one=**

8 T: **=+ohhk:ay+**
+nods head+

9 HAL: **=but some definitions are harder than the er**

10 **the [vocabulary items**

11 T: **[vo-vocabulary items=**

12 İLK: **OH OH**

13 GM: %**uh huh%**
%nods head%

14 HAL: **ı &saw a structure (.) damaging to end up**
&--1---> line 16

1-looks at item-writer group

15 **which ı don't know err: its [meaning**

16 SIB: **[what=**

17 HAL: **[°what does it what does it mean°**

18 BUS: **=&damaging to end up=**
&--->

19 HAL: **=ha- I am never never seen such a thing**

20 **ı googled it (.) [I- I couldn't find it**

21 T: **[↑damaging to end up existence of something=**

((reads the question silently))

22 HAL: **[YES unintelligible]**

(lines 20-29 omitted)

30 T: **er tanda yaptığımız gibi şey yapmamız lazım**

31 **discuss etmemiz lazım >çünkü bir definition**

32 **oraya da gidiyor buraya da gidiyor j ve h definitioni**

33 HAL: **=also: I couldn't find &adoring in the &book or**

34 **in the: curriculum °adoring° %is not(.)**
&--2--->
%--3--->line 25
2-heads up and looks at t
3- nods his head

35 T: **=zate:n %o definitionları Πnerden aldınız=**
besides where did you get those definitons
%---> Π--4--->
&--->
4:opens her hands to both sides and shakes her head)

36 HAL: **=cambridge Πdictionary or something**
Π--->

37 TUĞ: **=uh huh**

38 SS: **=uh huh**

39 SIB: **[°we have°-**

40 İLK: **[+some students+ unintelligible-**

+T looks at İlk+

41 T: **[E::VET**

42 SIB: **[°ama şey aslında burda adoring yazarken biliyordum**
but actually here while writing adoring I knew

43 **hani curriculumda olmadığını ama >hani good-looking var**
it is not included in the curriculum but you know there

44 **ortada ama (.)ama hani(.) ona da bi şekilde (0.2) er onla**
there is good looking but but you in a way with that

45 **bağdaştırıp (.) öyle öğrensinler [diye**
they can learn by connecting it in a way

46 TUĞ:→ **[ama face validity**
but

47 T: **[ama zaten**
but anyway

48 TUĞ: **[çöpe gitti bu sefer**
goes to the waste this time

49 T: **ama- ama zaten test ettiğin şey Ωgood loo:king**
but it is good looking that you are testing
Ω--4----Ω line 53
4: makes a fist with her left hand and raises it

50 **zor kısmın >good-looking olması lazım< (2.5)**
the difficult part must be good looking

51 HAL: **EVET**

52 T: **%hatırlayın (.) θdefinition onun bir seviye altında**
remember the definitions should be one level below that
θ--5--->
5: makes a fist with the right hand

53 **Ωoluθcak=**
Ω--->

54 HAL: **=θ°yes°**
θ--->

55 NUR: **±hmm ◊hmm±**
±nods her head±

◊take notes---> lines 58

Extract 3 begins with teachers signaling the shift to the next activity for the feedback session of group 5 and addressing preservice teachers to elicit participation. In the upcoming turn, Tuğ from feedback group 3 produces an (oh) in sotto voice while Hal from the other feedback provider group raises his hand and bids for a turn (Sahlström, 2002) and immediately after, receives a go-ahead response (söyle/say it) from the teacher. With a turn-initial (şimdi/now) and hesitation marker (er), Hal performs a turn-initial reference to a past learning (Can Daşkın, 2017) by explicitly taking the teacher as a reference point (you told us) for his forthcoming statement. In line 7, he refers to a testing principle (definitions should be an- an- minus one) as an account to preface his upcoming problematization, which paves the way for negotiation and evaluation of the focal test item in the following turns of the session. Also, in his formulation Hal demonstrates his knowledge regarding the rule for writing definitions which is taught by the teacher previously in the class as he makes a turn-initial reference to the teacher. In the next line, he signals his upcoming negative evaluation (Can, 2020) with the contrastive marker (but). He asserts that the definitions given in the vocabulary matching exercise are harder than the target vocabulary tested on the question. Before reaching the end of his TCU, the teacher performs an anticipatory completion (Lerner, 1996a) and completes his turn in a terminal overlap (vocabulary items), which shows her alignment and the noticing of the problem in the design of the question beforehand. In line 12, Tuğ from the other feedback group adds with the change of state token (oh) and gives an example from the target word given in the definitions box on the exam, which also displays his agreement and alignment with Hal's comment regarding the difficulty level. At the same time, Tuğ' comment also receives an embodied acknowledgment from his group member İlk. In line 14, Hal takes up his turn again to elaborate on his problematization by prefacing his turn (I saw a structure) and provides an example (damaging to end up) from the definition box provided in the question. Hal claims insufficient knowledge (I don't know its meaning) on the item and requests an explanation of the word through a wh-interrogative (what does it mean) from group 3. However, he does not receive a response to his request for information from any of the group members. In the meantime, Sib, the item-writer of the focal question, initiates

an open class repair with (what). In response, Bus sitting next to Hal, repeats the problematic definition and brings a resolution to Sib's repair while Hal sustains his gaze at Sib. Upon no response to his inquiry, he upgrades his negative assessment regarding the item by stating that he had never seen the definition before and couldn't find it even though he had searched for it beforehand. His statement indicates that he has done his job as a reviewer and checked the exam prepared by group 5 before coming to the class (Can, 2020). As a rule, established in the classroom in the first week of the summer school, preservice teachers must provide written feedback on each other's exams on the university's online platform (ODTUCLASS) before coming to the class. In the next turn, the teacher reads aloud the problematic definition, and Hal issues a confirmation token (YES) with a rising intonation. It is possible to say that the teacher might read aloud the problematized definition to check its grammaticality. In line 30, the teacher asserts that the item needs to be discussed in detail because of additional problems related to the other definitions. In the following turn, Hal sustains his gaze at the teacher, and with turn initial (also), he signals an upcoming additional comment. He problematizes another word, adoring, which is not included in the book and therefore presumed to be "untaught/uncovered" and therefore violates another testing principle "test what you teach." In the upcoming turn, the teacher shows her alignment through her head nodding and launches another TCU. She produces a request for information about the source of the vocabulary in the construction of definitions. At the same time, she opens her hands toward the item-writer group and sustains her gaze at Sib. Latching with the teacher's turn, Hal provides a candidate answer from a well-known source for the English language (cambridge dictionary or something) which receives a confirmation from the other preservice teachers through the token (uh huh) that displays their alignment with Hal. In line 39, Sib, as the item writer, starts formulating her response with inclusive language (we have) which indicates that she speaks as a representative of her group members. But she cannot continue and cuts off her turn, which overlaps with İlk's upcoming critique constructed through invoking non-present actors and students. Even though İlk's comment is not audible because of the overlap, the teacher turns to her and issues an agreement token at a faster pace. In line 42, SIB continues her yet -incomplete turn in sotto voice and starts reflecting on her item writing process (Can, 2020). While formulating her explanation, she accepts that the word adoring is not included in the

coursebook. However, she expected the students to make a connection with a semantically related word, good-looking, and match the words with one another. Her turn consisting of micro pauses and hesitation markers overlaps at the final utterance with Tuğ, who also launches a problematization with the contrastive marker (but). Tuğ asserts that by including a word that is not taught in the book, the item-writer violated the rules set by the face validity (ama face validity çöpe gitti bu Sefer/goes to the waste this time), which is in this case 'never test what you didn't teach' principle. Overlapping with Tuğ, the teacher starts her turn by repeating herself and states that the difficult part must be the target vocabulary 'good looking,' not adoring, which is given in the definition box of the question yet is not covered in the book. During the formulation of her turn, she articulates the token (zaten/anyway), which treats her explanation regarding the rules for constructing definitions as an assumed knowledge to the members of the session. In line 39, Hal shows his agreement and alignment with her by producing an agreement token (EVET/yes). In line 40, the teacher continues her multi-unit turn and provides a reference to a past learning (Can Daşkın, 2017; Can Daşkın & Hatipoğlu, 2019) about writing definitions (hatırlayın/remember) and treats the information she provided as studied in class as a rule. Besides, during her extended telling, she reminds the rule about constructing definitions in the form of advice delivered as "information." Her embodied gestures also align with her telling, which receives an agreement token from Hal in line 53 and a claim of understanding (hmm hmm) from a member of the item-writer group, NUR. In her embodied acknowledgment of the teacher's extended telling, Nur takes notes even though she did not construct the test-item. This displays her orientation to her role and responsibility (Heritage, 2013; Pomerantz, 1980) as the item-writer to be informed about the problems and necessary changes for the test item (Can, 2020).

It is seen in extract three that while actively orienting to their roles as feedback providers, preservice teachers identify potential problems in the test items through their employment of testing principles that are violated or must be fulfilled in the construction of valid and reliable language exams. While invoking the testing principles during their problematization sequences, preservice teachers, Hal, and Tuğ base their claims on common disciplinary knowledge and expertise in testing and

evaluation, which position them in (K+) position among their peers (Heritage, 2013). To promote the irrefutability of their claims, they bring an account to their claims through invoking rules as an interactional resource. It is also important that, while the preservice teachers provide accounts, they also demonstrate their learning regarding the testing rule. As they give a turn-initial reference to the teacher (you told us), it might indicate that they have added the rule into their second language assessment repertoire thanks to the teacher and use it as an interactional resource to back-up their negative feedback. Th Still, they also present examples from their own test-taking process. Preservice teachers pinpoint the problems and achieve mutual understanding on the issue by orienting to their experience while taking the test as a potential student (I couldn't find it, I saw). Besides preservice teachers, the teacher also invokes larger pedagogical principles (Waring, 2017), which are covered in class, as is evident in her use of reference to past learning events (Can Daşkın, 2017; Can Daşkın & Hatipoğlu, 2019) during her extended telling to deliver advice. During her formulation, the teacher also depersonalizes her advice by going general through her inclusion of all-preservice teachers (Schwab, 2011) to create a learning space during her advice delivery. It is evident in the item-writer group's note-taking process that preservice teachers collaboratively achieved a mutual understanding of the emergent problems that violate the testing principles by orienting to different interactional resources.

The following extracts 4a, 4b and 5 are analyzed to present evidence to a micro-moment of learning in the peer feedback interaction. Extract 4a is taken from the first feedback session of the summer school, in which group 3, consisting of TUĞ, İLK, USE, NEH, and OYA, received feedback for their grammar section. In the extract below, TUĞ, the item writer, receives feedback on the wrong placement of the instructions for the question from the teacher.

Extract4a. Week 4 – Group 3/ Instructions

1 T: **ok:ay so we have section <one in roman letters> (0.2)**

2 **then we have part a >which is not grammar questions**

3 **but just part a grammar< %right &we don't need a question**
%--1---> line 11
4 **in [unintelligible] part** &--2---> line 12

1:T looks at TUĞ

2: T nods head

5 TUĞ: **right yes**

6 T:→ **then we hav- you have to add #INSTRUCTI:ONS (1.5)**

7 **and then after that you say read the following questions**↑
#--3--->line 8

3:AYS take notes

8 **and choose the appropriate answer and mark on (0.5)**

9 **your (0.2)coding sheet to mark on is a transitive verb**

10 **so you have to mark on [your answer on or them#**
#--->

11 TUĞ: **mark then [on your sheet yeah**

12 T: **[on your coding sheet so you have to think about**

13 **rewriting that %part**
%--->

14 TUĞ: **hmm hmm&**
&--->

It is noteworthy that the teacher usually starts her turn after the peer feedback groups to summarize the main issues and give her final advice on test-item to eliminate any problems violating the reliability and validity of the language exams. Therefore, extract four begins with the teacher's follow-up turn after peer feedback groups commented on the grammar section of group 3. Between lines 1-4, the teacher reiterates a comment made by the groups and states that part A in the exam, which is not about the grammar, is redundant and, therefore, can be taken out.

Before completing her turn, she looks at Tuğ, who is the item-writer of the problematic part and consequently responsible for it. The teacher issues a confirmation request (right) from Tuğ, which arrives immediately with the token (right yes). In line

6, she continues her turn with the marker (then), which signals an upcoming additional comment from the teacher. In her statement, the teacher explicitly addresses Tuğ with the second person singular pronoun (you) and delivers her advice (you have to add the instructions) for the focal question. She further explains the construction of the instructions and also touches upon the ungrammatical item by reminding the grammar rule to fix the problem (mark is a transitive verb) and finally provides the grammatically correct version of the instruction. In the next line, Tuğ employs a full modified repeat (Stivers, 2005) at the final utterance of her turn which shows his agreement and alignment with the comment. During the teacher's construction of her advice related to the problematic question design, Ays, another member of group 3, takes notes on the exam paper they have prepared as a group until the teacher completes her TCU in line 13. On the other hand, Tuğ produces an acknowledgment token (hmm hmm) in coordination with head nodding.

In extract 4a, to recap the feedback provided earlier in the interaction, the teacher summarizes the problematic parts identified in the test item while, on the other hand delivering her advice by pointing to a face validity problem which is the wrong placement of the instructions in the question stem that can create possible problems during the test-taking process of 6th graders. According to Heaton (1990), "instructions should come first because they will lead the students to answer the questions according to the context." Therefore, the teacher's pinpointing the violation of a face validity rule covered in the first weeks of the summer class invokes a rule-policing (Balaman, 2016) which is the "placement of the instructions before the question."

In terms of the uptake of the advice, while we don't know whether there was a change in Tuğ's epistemic status after the teacher's feedback, we can say that his group member, Ays took notes during the delivery of the feedback, which shows her noticing of the problem after the negotiation held for the grammar parts A and B. According to Heritage (2013) and Can (2020), as an item-writer in group 3, taking notes during the advice-delivery of the teacher displays that Ays shows her responsibility and role as the test-writer to know and mark the changes offered to the test item.

According to Waring (2015, 2017), recipients might display their uptake of the advice differently after the advice delivery during the teacher-mentor feedback sessions. First, through the simple acceptance tokens such as okay, yes, as Tuğ employs during the teacher's advice-telling or with nonverbal actions such as head nodding and note-taking, which are both present among the item-writer group members' bodily orientation to the feedback. Besides, as problems "trigger the delivery of advice" (Waring, 2017), the teacher presents her advice step-wise by expressing the steps that must be taken to construct natural, grammatically correct, and useful instructions to guide students on the exam. This further brings evidence to teachers' use of advice-giving as an opportunity to increase learning opportunities in class through the inclusion of all preservice teachers in the classroom (Schwab, 2011).

In extract 4b, Peer feedback group 4, consisting of HAL, BUS, ZEK, and GUL, delivers feedback to TUĞ as the item writer on the wrong placement of instructions again in another question, question 10 in the same session dedicated to grammar sections of the language exams.

1 T: **hay:di bakalım question ↑ten**
okay let's go question ten

2 TUĞ: **o::h >no no no no<**

3 SS: **uh huh**

4 HAL:→**&şe:y yukarı mı gitmeliydi acaba instruction(.) yani soru=**
should it be at the top instruction I mean the question

&--1---> line 6

1:HAL looks at teacher

5 T: **=evet kesinlikle #ilk önce soru ↑gelmesi lazım >siz &nerde**
yes exactly first the question should come where did you
#--2--->line 9
2:AYS TAKE NOTES
&--->

6 **gördünüz ya< belki çok gördüğünüz için ama >ilk önce SO:RU**

see maybe because you have seen it a lot but the question

7 **[%gelicek**
should come first

%--3---> line 11

3:TUĞ nods his head

At the beginning of extract 4b, the teacher nominates question ten as the focal point in the discussion. Tuğ who is the item-writer of the multiple-choice questions in the grammar section, expresses his humor with the polarity marker (no no no no), which receives a shared laughter token from the other preservice teachers in the session. In line 4, Hal self-selects with the turn initial hesitation marker(şey/well) and suggests replacing the instruction on the top of the page (şey yukarı mı gitmeliydi acaba instruction yani soru)in the first language. According to Üstünel and Seedhouse (2005), among many uses of code-switching, teachers might use their first language to 'give feedback, which is a delicate issue in terms of face issues (Goffman, 1955) that might arise and interfere with relational issues between the parties.

At the same time, while formulating his suggestion with his orientation to a testing principle, Hal frames his suggestion in the form of a confirmation request (gitmeliydi acaba) which presents the suggestion as an offer for other preservice teachers to negotiate and confirm its effectiveness collaboratively and appropriates on the problematic test item (Can, 2020). At the same time, Hal epistemically downgrades his epistemic priority as being the first person to suggest a change on the item (Heritage & Raymond, 2005). Besides, Hal's suggestion in the same feedback session of group 3 is an act of 'displaying knowledge' since the rule concerning the placement of the instructions has been presented as advice in the teacher's extended telling for a previous problematic test item in the previous discussion of Tuğ's grammar section. In line 4, the teacher shows her alignment and agreement (evet) with HAL's comment by approving him with the certainty marker (kesinlikle/exactly). Between lines 5-7, in her multi-unit TCU, she elaborates on Hal's comment in relation to the placement of instruction. She repeats her exact words from extract 4a to remind the rules for the instruction as a part of face validity. During the formulation of her

advice, she produces a wh-interrogatives and asks the item-writer group about where they have seen the placement of the instruction after the question. However, without waiting for an answer to her question, she makes an inference about its wrong placement by attributing it to preservice teachers' prior testing experiences. Therefore, it can be inferred that teacher's question was not a real information-seeking question (Koshik, 2011).

On the contrary, it is a means of cementing the understanding of the severity of the problem, which is already established as problematic. In her termination of his TCU, the teacher receives minimal response tokens from the exam and feedback provider groups (hmm hmm) and signals the shift to the next activity.

Apart from the previous extract, where we have seen the teacher's advice-giving for a problem in the test item by invoking testing principles, in extract 4b, we see how a preservice teacher from the feedback provider group displays his understanding of the teacher's advice by invoking the same principle in the evaluation of another test-item in the same grammar section of the item-writer group. We can say that the peer feedback group treats the wrong placement of the instruction in the question stem as problematic and formulates his problematization in the form of advice to change the place of the instruction by invoking the same principle which is stated previously by the teacher in the same feedback session.

Extract 5 is taken from the fifth session in which preservice teachers work on vocabulary sections of the language exams. In the extract below, TUĞ and ILK present their self-feedback to the classroom before the start of the peer feedback session. As a rule, established by a joint resolution in class, the spokesperson from each group takes the floor to talk about the problematic parts they are aware of in their exams at the beginning of each session, before the official peer feedback session begins. In the extract below, after İLK completes her TCU, TUĞ takes the turn and starts talking about some problematic parts on their exams before their feedback session begins.

Extract 5. Week 5- Group 3 / Instructions

1 TUĞ: **>so don't worry about that we realized the issue**

2 **and we will make sure to have everything back in the**

3 **correct order**

4 T: **+hmm hmm+**
+T nods head+

5 TUĞ: **+AND: ı ı also like to mention tha:t >in the vocabulary**

6 **part< ı think they are supposed to be an alphabetic order**

7 **(0.2) &am ı correct&**
+T looks at TUĞ--->line 11

&T nods head&

8 T: **they are they are hmm hmm**

9 TUĞ:→ **%A:ND ALSO↑ >the instruction should #come first< (.)right#?**
%turns to teacher--->line10 #raises his hand #

10 T: **HA::H ↑ (0.2) YOU A:RE LEARNING**

((opens her hand on both hands and leans on TUĞ))

11 TUĞ: **I KNO:W**

12 T: **%UH HUH+**
%---> +--->

In extract 5, Tuğ reflects on the written feedback they have received from the peer feedback groups on the university's online platform (ODTUCLASS). At the beginning of the extract, Tuğ clarifies the problematic page design of their exams resulting from the Pdf and word versions. He replies to the feedback delivered as advice by the peer groups by employing a complex advice acceptance which is usually formulated as "I will do X" (we will make sure to have…). In the following line, with turn, the initial discourse marker (and also) Tuğ signals an upcoming additional comment for the vocabulary section of their exam. During the formulation of his turn, he starts talking about rules related to the design of the test items and distractors. He states that

the distractors of the multiple-choice questions should be alphabetically ordered, which is another principle they have covered multiple times in the previous sessions. While bringing the violated principle to the interaction, he performs a kind of a self-policing (Balaman, 2016). In his turn, which is epistemically downgraded with the stance marker (I think) and (supposed to be), he requests confirmation (am I correct) from the teacher by sustaining his gaze at her. His request from the teacher indicates that he accepts the teacher as an authority in the classroom and therefore resorts to her expertise and disciplinary knowledge in testing and evaluation. In the next turn, the teacher confirms him and adds that they are already alphabetically ordered, which does not treat the issue as problematic. In line 9, TUĞ proceeds with his turn again with a rising intonation (and ALSO) to state that instructions should come first before the question, which was an issue that he has received multiple critiques in the previous sessions for the grammar section that he designed in week 4. Although it is presented in extracts 4a and 4b that he only produced minimum response tokens such as (hmm hmm okay) to the advice, his restatement of the issue as self-feedback to his vocabulary section in one session later shows that his epistemic status changed from (K-) to (K+). In the formulation of his turn, he seeks confirmation again (right) from the teacher who articulates a change of state token (ha:h/oh) (Heritage, 1984). It is important to note that (K-) participants seek elaboration on the matter while (K+) participants seek confirmation. Tuğ's invoking of the rule regarding the placement of instruction is formulated as a confirmation-seeking statement which fortifies his (K+) position.

In the next turn, the teacher treats Tuğ's comment as learning by explicitly stating (you a:re-learning) regarding the placement of instructions as a rule covered in the face validity section during the first weeks of the summer school as well as in the first feedback session, week 4. According to Seedhouse (2008, 2010), "learning involves a change in the cognitive state" of interactants by adopting a linguistic item or a pattern that is treated as "missing" in the first phase but acquired in later stages through the scaffolding of the teacher. Therefore, during the feedback session in week 4, Tuğ's placement of the instruction after the question was treated as problematic, which placed him in the (K-) status. However, after receiving feedback from his peers as well as the teacher, his epistemic status changed from (K-) to (K-), which displays his

epistemic progression (Gardner, 2007). According to Markee (2008) and Doehler and Lauzon (2015), Tuğ's ability to incorporate an item into his interactional repertoire in language assessment across different interactional events throughout the sessions displays how he made an epistemic progress thanks to the peer feedback interaction. If we examine the emergence of a change in the epistemic status of Tuğ with the conceptualization of Seedhouse's (2010) learning model, we can say that;

1-In the precious moments of the interaction, the absence of the testing rule led to the wrong placement of the instruction in the question stem, which is treated as problematic by the peer feedback groups and the teacher. Therefore, a collaborative orientation to repair the trouble is initiated by the teacher through his extended advice delivery packed as information. In the second turn, the problematic issue unfolded in the same session, which is reminded by one of the peer feedback groups through another formulation of a suggestion and received another feedback from the teacher in her follow-up turn about the same problem.

2-During these interactional events, Tuğ's cognitive state displayed in his verbal `(okay hm hmm)` and nonverbal behavior (gaze, note taking) in and through interaction throughout the delivery of feedback by the teacher and the peer feedback groups.

3-In the last phase, his self-rule policing through the testing principle concerning the placement of instructions in the question stem at a different point in the interaction shows "a micro-moment of potential learning as observable" (Seedhouse, 2010), which is also treated as *learning* by the teacher at the moment it became recognizable in the interaction (Seedhouse & Walsh in Seedhouse, 2010, p. 128)

All things considered, extracts 4 and 5 display an example of "reference to a testing principle," that is, the placement of instructions before the question stem (Heaton, 1990) which emerges as a rule breach and potential problem in the construction of the test-item is eliminated by collaborative negotiation among the present parties in the interaction.

To start with, the teacher treats the "wrong placement of the instruction" as a problem and packs her advice in the form of information delivery by referencing the testing

principle the problem originates from. Later, one of the peer feedback groups, Hal, faces the same problem and also issues a suggestion by employing another reference to the same testing principle. In the last phase, the same testing principle was brought into the interaction in a new context (extract 5) by Tuğ who was the addressee of the suggestions provided in extracts 4a and 4a, which displays his uptake of the suggestions and also implementation of the rule. Considering everything, we can say that referencing a violated testing principle during the feedback session functions as a bridge between the missing and new information and fortifies the delivery of feedback to treat the emergent problem, as is evident in Tuğ's change in his learning state. Suffice to say, referring the testing principles during the peer feedback sessions contributes to the change in the preservice teachers' learning state (extract 5).

4.2. Employment of RTP in Suggestion Phase

As it is presented in the previous extracts, the identification of a potential problem makes the next relevant action a suggestion phase. In this phase, preservice teachers and the teacher collaboratively work on the problematic items to bring a possible solution. Therefore, the next extracts are analyzed to display the emergence of RTP in suggestion phases. In this section, Extracts 6a and 6b are analyzed to indicate how preservice teachers orient to the testing principles 1) when they raise counterarguments to candidate suggestions violating testing principles and 2) to back up their claims when they suggest a change on the problematic item.

Extract 6a is taken from the last feedback session of the summer school, which is dedicated to the writing sections of the exams. This extract displays how preservice teachers orient to the testing rules while critically evaluating the suggested changes for the problematized items in different sections of the exams. In the previously analyzed extract 3, Group 3, consisting of TUĞ, İLK, UZE, NEH, and OYA, received negative feedback from peer feedback groups on writing problematic instructions. Their instructions in the question stem are general, vague, and complicated, therefore yields ‘more than one possible answer 'problem in the item. In extract 2, Hal problematized the instructions in terms of two testing principles: 1) instructions’ are written in a difficult language 2) leading to more than one correct answer. After the mutual

recognition of the problem, the teacher initiated the suggestion phase of the interaction given below to overcome the problem in the question stem, which aims to test prepositions and numbers in the writing question designed for 6thgraders by the item-writer, Neh. In the extract below, Tuğ as the item writer presents a suggestion to the focal item which is challenged by Has and the teacher collaboratively.

Extract 6a. Week 7- Group 3/ Options

1 T: **#ne ∂yap↑SIN (0.4) <hadi sınıfça (0.2) NEH'e birazcık**
what does she do let's as a class just a little

2 **yardımcı olalım (0.4)**
help Nehir

#--1---> line 3

∂--2---> line 4

1:HAL raises his hand

2:T fixes her gaze at class

3 T: **artık siz bu kadar çok han:i (0.2) testing hani konuşan**
Now you are this much you know talking about testing

4 **insanlar olarak mesela <gruptaki #diğer arkadaşları↑**
as people for example her group friends
#--->#

5 **ne yapsın mesela>[nehir yani ne ∂yaparsınız**
∂--->
what does she do I mean nehir what do you do

6 TUĞ: **[ı will s-**

7 T: **[↑yani ne yaparsınız=**
I mean what would you

8 TUĞ: **=ı will sug~gest**

((clears his throat))

9 **uh huh excuse me ↑ ı would suggest to: go with SA::Y(0.3)**

10 **what ΛSELECT (0.2) three items (0.2) select three app-**

Λ--3--->line 12

3:Holds the exam paper and marks the section B

11 **appliances and then select >one two three yes three<**

12 **prepositions and FO:RM three full sentences (.) full**

13 **stop (0.5)**

14 T: **hmm hmm hmm hmm**

15 TUĞ: **Λfor B↑ °at leatst°=**
Λ--->

16 T: **=FOR B**

17 TUĞ: **FOR A↑:: ı would use the same structure aGAIN (0.2)**

18 **I would say (0.3) >select perhaps< FO:UR (0.2) this time**

19 **I don't know pardon four items >from the list a<bo::ve**

20 T: **hmm hmm+**
--->+

21 TUĞ **:A:ND(0.3) write +full sentences (0.2)**

22 T: **+oh**

23 TUĞ **:<using wor[ds=**

24 **[°=which words°**

25 TUĞ: **[not numbers [words**

26 T: **[to writ- to wri˜te (.) th-their prices=**

27 TUĞ: **=[YES**

(0.3)

28 T: +**ohhkay+**

+nods head+

29 TUĞ: **[>does that make [sense<**

30 HAS:: **[βHOCAM**

β--4---> lines 32

4:T tuns to hasret

31 HAS: **°may I say something°**

32 T: **sure >tabii ki**
of course

33 HAS:→ **<but what if they choo::se fo:ur >out of< th- (0.2)**

34 **six of them ↓here=**

35 T: **=hmm hmm**

36 HAS: **yani >they will choose the ones they k:now<**
---->β

37 T: **hmm hmm**

38 HAS: **so we neve- we will never learn (.) if they know**β

39 **↓the others=**
β--->

40 T:→ **[=opti:on vermiyorduk hani?]**
remember that we don't provide any options

41 HAL: **[°şey altıncı sınıf dimi bunlar=°]**
er they are sixth graders right

42 BUS: **=°şey altıncı sınıf aynen°**
they are sixth graders yeah

43 TUĞ: **hm:::**

44 T: **$uh huh$**

45 TUĞ: &**T:RUE**
&nods head---> line 49

46 T: **°o zaman°**

47 TUĞ: **ya [hep↑si**
either all or three items of three of them

48 T: **[ya hiç**

49 TUĞ: **>ya da üç tane ve üçü<**

50 T **+hmm hmm °aynen öyle°+**
exacly
+-------T nods--------+

51 TUĞ: **u:h hmm &hmm**
--->&

Extract 6a begins with the teacher's invitation of preservice teachers to bring a possible solution to the problem identified in question 9, prepared by Neh from item-writer group 3. While inviting preservice teachers to an advice-giving environment, the teacher addresses the whole classroom in the first language. According to Üstünel and Seedhouse, (2005), switching to a first language in the classroom might have a variety of functions. The teacher's use of first language here might originate from her intention to increase participation in the session. In the following line, T addresses the whole classroom (bu kadar çok testing konuşan insanlar olarak artık) as a

community experienced in testing and evaluation, which signals her expectation of preservice teachers to offer possible solutions to the problematized item. In line 4, the teacher also addresses group 3, which is the item-writer group and hence responsible for the focal item besides Neh. Her inclusion of test-item writer groups and the whole classroom displays her creation of a multilogue interaction (Schwab, 2011) and advice-giving atmosphere in the session. While addressing the classroom, Hal from peer feedback group 4 raises his hand; however, the teacher does not orient to his bidding for a turn (Sahlström, 2002). Therefore, he lowers his hand when the teacher shows bodily orientation to group 3 as the item-writer group. In a terminal overlap with the teacher's final TCU in line 6, TUĞ self-selects to provide a possible solution to the focal test-item and switches back to English. He secures his extended telling between the lines 6-19 with the continuation `(and)` and listing markers `(then)` and suggests selecting three items and three prepositions from a list to form whole sentences in the B section of the question. With no time-lapse, the teacher aligns `(hmm hmm)` and nods her head multiple times during TUĞ's formulation of his suggestion which can signal her affirmation in an embodied way. In line 15, Tuğ adds that his suggestion is for the B section of the exam and shows section B by pointing with his deictic gesture (Mondada, 2007) on the exam paper.

In the following turn, T produces a confirmatory repeat (You, 2014) and affirms that the suggestion is for section B of the question. With no time lapse, Tuğ takes up his turn again in line 14 and continues his suggestion to select four items from a list of prepositions to write prices in words for part A of the question, which is highlighted with the uncertainty marker `(I don't know)` and delivered with the advice-structure 'I would do X' (Shaw et al., 2015) that shows his side of the telling by explicitly referring to himself. During his formulation of the suggestion for part A, the teacher acknowledges him (hmm hmm) with an embodied head nodding and issues a clarification request `(which words)` in line 24 to Tuğ's `(full sentences using words)` utterance. In an overlap between lines 24 and 25, Tuğ adds that he aims to use words instead of numbers `(not numbers words)` under the pictures in the item. However, as Tuğ did not specify the aim of using words instead of numbers, the teacher issues another clarification to his suggestion (to write their prices) which immediately receives a confirmation `(yes)` from Tuğ in line 27 and a successive

acknowledgment (okay) from the teacher. In line 29, Tuğ produces a comprehension check to receive a response from his group members who did not show any orientation to his suggestions, but his turn overlaps with Has's addressing the teacher and asking for permission to speak, which leaves Tuğ's comprehension check unattended. In line 32, the teacher produces a go-ahead (Schegloff, 2007) response (sure tabii ki) by code-switching. In the following line, Has articulates the contrastive marker (but) as a preface to signal her upcoming objection (Can, 2020; Park, 2014) to Tuğ's suggestions in the previous turn. She employs (what if) hypothetical conditional (Linnel, 2002) in tandem with the third person plural pronoun (they), which invokes non-present actors (Leyland, 2021; Yöney, 2021) in this case, potential students as the test-takers in her objection to TUĞ's suggestion. She further adds that according to Tuğ's question design, the students will likely choose the prepositions they know, thanks to the extra options provided in the stem (what if they choose the ones they know so we will never know if they know the others). In her objection, packed as an interrogative to challenge the suggestion offered by Tuğ, Has brings a possible student behavior to the immediate environment and problematizes the previously offered suggestion and its possible future result (we will never know if they know the others) by going general with the use of inclusive language (we) which depersonalizes the problematization and avoids using the second person pronoun (you) to address the test-writer group in line 38. During the formulation of her objection, Has implicitly shows orientation to the violated testing principles 'avoid using excessive and ineffective distractors to eliminate guessing factors in the construction of options by referring to the provided distractors with the pronoun (the ones) and noun (the others).

To pinpoint the additional problem resulting from the violation of another principle in the suggested design of the item, Has displays orientation to the violated testing principle and invokes non-present actors in her formulation of the objection. In her invoking non-present actors (they she places the students as test-takers (Yöney, 2021) and puts the whole classroom (we) in the category of test-makers which creates a standardized relational pair (Leyland, 2021; You, 2015). During her turn in progress, the teacher shows her alignment with an acknowledgment token (hmm hmm) and nods her head. In her follow-up turn in 40, she smiles and bodily orients to TUĞ and shows

her agreement and affiliation to Has's comment by employing a past reference (Can Daşkın, 2017; Can Daşkın & Hatipoğlu, 2019) to the violated testing principle (option vermiyorduk hani) which explicitly refers to the violated principle Has oriented in the formulation of her objection. During her turn in progress, the teacher reminds TUĞ that they aren't supposed to provide extra options to students as a rule which they have learned either during the lecture sessions of the ELTE course or from their assigned readings in testing & evaluation, as she displays with her choice of language inclusive (we) and past tense (vermiyorduk hani).

In line 43, Tuğ responds with the elongated token (hmm) while the teacher produces a laughter token, and in line 45, Tuğ issues a strong confirmation token (true). In her turn-entry in line 46 teacher directly asks Tuğ about what can be done (o zaman/then) in the first language after mutually agreeing on the problem in his previous suggestion. In line 47, Tuğ produces an either-or structure (ya hepsi/either all) which receives an anticipatory completion (Lerner, 1996a) from the teacher (ya hiç/or none) in an overlap in line 48. In the next turn, Tuğ continues with the token (ya da /or) and reformulates his previous suggestion in line with the teachers'prompt, and offers to structure the options by giving three items from the list of prepositions (ya üçü ya da üç tane/either tree or three of them). His reformulation of the new suggestion shows his understanding of the design of the options as he displays his understanding (Koole, 2010) and hence agrees with the comments provided by his peer and the teacher. In line 50, the teacher immediately shows her affiliation with the acknowledgment token (hmm hmm) and strong agreement marker (aynen öyle/exactly), which receives response tokens (uh::, hmm hmm) from TUĞ in line 51.

Apart from the previous extracts analyzed in the problematization phase of the feedback session, extract 6a depicts how preservice teachers in collaboration with the teacher orient to the testing rules while critically evaluating the suggested changes for the problematized items in different sections of the exams. It has been shown that preservice teachers employ a wide range of principles in different phases of the interaction. However, their orientation aims to fulfill the common objective of the course, to provide quality feedback to the exams under review to ensure the validity

and reliability of the language exams. In line with this aim, preservice teachers share the responsibility and collaboratively work on the decision-making process for the suggested changes on the problematic items. As a member of the item-writer group, Tuğ brings a possible solution to the problematized item, which Has further problematized when she notices a potential problem in his suggestion. Has's orientation to the violated principle by referring to non-present actors, students further receive alignment from the teacher. She reminds the violated principle and paves the way for an epistemic change in Tuğ's epistemic stance, as it is further evident in his response token and reformulation of his previous suggestions to display his understanding. Has's implicit orientation to the testing principle further receives affiliation from the teacher in her follow-up turn when she explicitly refers to the principle 'window dressing, that is, avoid excessive verbiage and irrelevant clues leading to the correct answer.' (Haladyna, 2004). Therefore, we can say that the collaborative orientation of the preservice teacher and the professor to invoke a violated principle in challenging a suggestion leads to the discovery of a better alternative for the problematized item in the writing section of group 3. The further negotiation process for a new suggestion with the initiation of Hal is presented in the following extract 6b.

Extract 6b. Guided

((HAL raises his hand T turns to halil))

1 HAL: **can we (.) put (.) li:ke=**

2 T: **=+hmm hmm**

--1---> line 8

1:T nods head

3 HAL:→**we we give (.) a guide:line**

4 T: **hmm hmm**

5 HAL: **(0.2) er for example refrigerator COSTS >bla bla bla**

6 **(0.2) and then we want the >number <t:here (0.3)**

7 T: **hmm hmm**

8 HAL:→**TH:AT'S guided +[like-**
+--->

9 TUĞ: **[°it depends on the example=]**

10 T: **[=IT DEPENDS ON ϑWHAT NEHİR WANTS**
ϑ--1---> lines 69

1:T turns to Nehir

11 HAL: **oh YEAH**

12 T: **↑what is your OBJECTİ:VE**

13 HAL: **just suggestion ı am**

14 T: **[DO YOU: want u:s (0.2) OR do you want to check whether**

15 **>the students are able to construct grammatically correct**

16 **senten(.)ce:s using the words and the numbers(.)< or DO::**

17 **Y:OU want to ON:LY %check >whether% the students know↑how**
%halil nods his head%

18 **to write the numbers< ↓in words=** (0.3)

19 NEH **=°no°**

20 T: **=what is yo:ur AIM=**

21 NEH: **=er: numbers (0.2) a:nd er: prepositions(0.5)**

22 T: **so >↑just numbers<**

23 NEH: **hmm hmm=**

24 T: **=thenϕ YOUR &SUGGESTION IS::&plausibleϕ**
ϕ-------T points to HAL--------ϕ
&Hal nods head&

25 **SO you can s::ay the refrigerator I::S right(0.2) and then**

26 **in parenthe:sis you give the price of the refrigerator (.)**

27 **and then you instruct students >to write the numbers they<**

28 **↑see:: in whhords(0.2) it's much EASIER but the format of**

29 **the question no:w changes**

((ELA sneezes))

30 T: **çok yaşa ohkk::ay**

31 NEH: **°hm hm° okay=**

32 T: **=does that fulfill th-the objective (.) on your minds**

33 NEH: **yes ohkk:ay**

34 T: **then we structured (.) the question one (0.3) now what**

35 **about question two**

Before the extract, Hal, from the feedback-provider group, raises his hand and bids for a turn (Sahlström, 2002). Once he establishes a mutual gaze with the teacher, he initiates his suggestion to Nehir's previous problem, which is an unstructured, general and vague question stem that leads to more than one possible answer in the writing section of group 3. Hal signals an upcoming suggestion in the form of a question (can we put like) which downgrades his statement with the uncertainty marker (like) and the modal verb (can). More importantly, he formulates his suggestion as an example for preservice teachers to consider and approve its appropriacy (Can, 2020), as is evident in his use of inclusive language (we).

Without a time-lapse, he receives a verbal acknowledgment token (hmm hmm) and nonverbal nodding from the teacher and continues his formulation in the following line. At the beginning of his turn, he self-repairs himself with the repetition of inclusive (we- we) and changes the structure of his formulation from the question format to the declarative (we give a guideline), which provides an account of his upcoming suggestion in line 5. In what follows, the teacher responds with an acknowledgment token and nods her head, indicating her alignment to HAL's turn in progress.

In the following line, Hal formulates his suggestion as an example with the discourse marker (for example) and reads the new version of the question stem (refrigerator costs bla bla) out loud. In his suggestion, he changes the problematic question stem by providing the item, refrigerator, as an incomplete sentence and requesting students to write the price in the blank spaces with the numbers (and then we want the number there). After a brief gap of silence, he adds the demonstrative (that's guided like), which functions as a post account (Waring, 2007) to his suggestion and refers to the testing principle which he aims to

fulfill “writing tasks must be structured with guiding instructions and a series of pictures to restrict and guide students to write a specific form of task (Davidson & Lynch, 2002, 2008; Heaton, 1989)”.

Instead of a full reference to the principles directly as in the previous extracts of problematization, preservice teachers employ a fine-tuned use of testing principles while formulating their suggestions in the last weeks of the feedback sessions. This might arise from the fact that preservice teachers might have added the principles of language testing into their language assessment literacy; therefore, they show their orientation to these principles with fine-tuned language use in the later weeks of the sessions. Besides, while referencing this principle, Hal shows his orientation to the violated testing principle, which is the unstructured, unguided instruction in the question item that he problematized in extract three by expressing the problem from his side of telling. Therefore, we can say that his fine-tuned use of principle originates from his earlier problematization of the same test item; consequently, he knows the violated rule in constructing the item he refers to while offering his suggestion.

In what follows, in a terminal overlap in the final syllable of his word, Tuğ, who is also a member of the item-writer group, comments on Hal's suggestion with the phrasal verb `(it depends on the example)`, which receives a partial modified repeat (Kim, 2002; Stivers, 2005) from the teacher in 10 by using his turn initial phrasal verb `(it depends on what nazlı wants)`. Aligning with Tuğ's comment in her turn, the teacher emphasizes it depends on Neh, the test-writer, as she is the one constructing the item and has the epistemic right and responsibility (Heritage, 1984) to know and evaluate the appropriacy of the suggestion being offered to the item. In line 11, Hal provides a change of state token `(oh)` and an acknowledgment token `(yeah)` to the teacher's comment. In the following line, the teacher bodily orients to Neh and issues a wh-interrogative `(what is your objective)` in line 12. She puts Neh as the item writer in (K+) position by producing a question related to the item writer's epistemic domain because she has constructed the question according to the demands of the specific teaching context the exam aims to answer.

At the same time, in line 13, HAL self-selects and weakens the strength of his suggestion with the hedge (just suggesting ı am). However, since Neh does not immediately orient to the teacher's question, the teacher elaborates on her question between lines 14-18. She produces an alternate question (Koshik, 2002) with (or) and asks Neh what she aims to test, whether to learn students' ability to form correct grammatical structures while writing prices of the items given in pictures or the ability to use only numbers. Upon a brief gap of silence, Neh responds with the negative response token (no) in a sotto voice. In line 20, the teacher repeats her question in line 12 (what is your objective), which indicates that she found the response insufficient and asks Neh's aim again to be able to decide on the suggested change. In most advice-giving interactions, wh- interrogatives function as pre-advice or pre-proposal sequences (Hepburn & Potter, 2011) directed at the advice-recipient before issuing advice that best suits the recipient's interest. However, in this extract, it comes after the suggested change to check its appropriacy to the item writer's objectives as she is the one who created the test item and holds the right (Heritage, 1984) to judge its appropriacy.

Upon latching with the teacher's repetition of the question, Neh responds by stating that the test aims to measure numbers and prepositions as provided in the exam specifications. In her formulation of the answer, her turn includes multiple micro pauses (0.2) and elongated hesitation markers (er:). In the following line, the teacher produces a full modified repeat of Neh's previous turn (so just numbers) with the rising intonation, which signals her confirmation request. In line 23, Neh acknowledges (hmm hmm) the teacher's request. Her acknowledgment immediately results in a change in the teacher's bodily behavior. She turns back and points with a deictic gesture (Mondada, 2007) to Hal, who announced a suggestion at the beginning of the interaction. While HAL responds with a nonverbal head nod, the teacher signals her continuation with the marker (↑so) in rising intonation and elaborates on the suggestion Hal provided before in her extended telling between lines 26-29. She shows her affiliation and agreement with the earlier suggestion by employing a full modified repeat (Kim, 2002; Stivers, 2005) of Hal's suggestion in line 5 (you can say refrigerator IS:) and signals the ending of her TCU by summarizing the format change in the focal question. She checks understanding with the token (okay) and

between lines 31 and 33, Neh claims understanding with the acknowledgment tokens (hmm, hmm, yes, okay). According to Waring (2007), advice receivers show their acceptance of the advice provided by the teacher in two different ways: Firstly, by engaging in note taking and or with minimal tokens such as okay, yes, hm hmm, they can display simple acceptance of the advice. Secondly, they can show complex acceptance of the advice through formulations such as 'I will do X' (Swaw, et al., 2015). Therefore, we only see a simple acceptance of Neh's advice through her articulating the tokens (yes, okay) respectively. However, we only see her response after the teacher's orientation to her, which might display that item-writer groups may take the teacher as an epistemic authority in the classroom during their decision-making process.

In contrast with the extract 6a, where we have seen the cancellation of a suggestion that violates a testing principle, in 6b, it has been demonstrated that preservice teachers also intend to employ testing principles to back up their claims when suggesting a possible solution to the problematic test-item. In extract 6a, a suggestion offered by another item writer in the same group has been canceled based on a violation of a testing principle in the design of the options. In extract 6b, the second suggestion has been accepted directly without further negotiation as it is presented in the directive format with a post account based on another principle to be fulfilled to ensure its appropriacy to the question stem. The alternative suggestion offered to the problematic item immediately receives a response from another member of the group and the teacher to ensure its appropriacy to the interests of the item writer. It directly gets accepted as a solution to the problematic item. However, as the item writer group consists of another four preservice teachers, we cannot claim if they display mutual consensus as a group on the decided change. Yet, Neh as the item writer, and Tuğ as the other group member, take notes during the interaction; we can infer that they have displayed a nonverbally display of understanding and acceptance of the advice.

Extract 7 is taken from the first session dedicated to the grammar sections of the language tests. In the extract below, group 1 consisting of HAS, REY, BAR, and OZG receive comments on the problematic test item from group 4, HAL, BUS, ZEK, GUL.

This extract is chosen for presenting the engagement of one of the item-writer preservice teacher in the suggestion phase of their language test.

Extract.7-Group 1/ odd one out

1 T: **oka:y but this should be revised (3.4)question seven (3.1)**

2 ELA: **°we have already discussed it=°**

3 T: **=we have already discussed it but let's focus on what is**

4 **tested a:nd ↓how it is °tested=**

5 HAL: **=wh questions [°unintelligible**

6 T: **[↑SO:: ho:w(.) when(.) which(.) how often(0.2) ↑right=**

7 HAL: **=[yeah**

8 ELA: **[°yeah°**
(0.3)

9 T: **does he go to his work comma (0.2) o:h uhm I am sorry**

10 **does does >does he do his homework at eight o'clock<**
(0.3)

11 HAL: **↑%may:be er how often can be replaced with ↑where (0.2)**
%--1---> line 14

1: Hal looks at teacher

12 T: **hmm hmm**

13 HAL→ **because it's like a bit longer than the other options and**

14 **it's not a correct answer %so:: °it a:sks ↓wh-ques[tions°**
%--->

15 T: **[ohhka:::y how(.) when(.) which (.) ↑a:::nd=**

16 HAL: **=+where [mesela+**
+-----T nods----+

17 T: **[↑ A:ND this is a good suggestion because ♥you have (.)**
♥--2---> line 25

18 **how often already tested he:re in δquestion six**

```
                                                    δ--3--->
      1:Has nods and writes on the exam paper

      2:Bar nods and takes notes
19          now you have to find something ELSE  ↑right IF you have

20          [WHE:RE↑

21    REY:  =maybe we can just put wh- forms li:ke what(.) where(.)

22          [when(.)
            (0.2)

23    T:    [yeah (0.2) δbut using how often aga:IN it's (0.2)over

24          +testing it °in a +way°  right
                  --->δ
            +BAR nods her head+

25    REY:  =♥&okay&
              &REY nods&
              ♥--->

26    T:    ↑RIGHT >you have already tested whether the students

27           kno:w< (.)the meaning of how +often in question six+

                                              +--Bar nods her head--+
28          and now you are testing it &again here ↓in question
                                        &HAS nods--->
29          ♥seven >you are testing the same thing over and over

30           &again< ↓okay
             &--->
31    HAS:  °okay°

32    T:    now (.) question eight
```

Extract 7 begins with the teacher nominating question 7 as the focal point in the feedback session. After a long pause of 3.1 seconds, in line 2, Ela self-selects and informs the teacher that question 6 is already being discussed in class with inclusive language (we have already discussed). In line 3, T aligns with her comment by

employing a full modified repeat (Kim, 2002; Stivers, 2005) of Ela's previous turn. Yet, in the same turn, she employs the contrastive discourse marker (but), which signals an upcoming disagreement to pursue the same topic offered at the beginning of her turn. (let's focus on). Even though the item has been defined as problematic on "overtesting"', the teacher prefers to discuss the aim and structuring of the item in detail, as they have done for each test item until now. Latching with T's turn, Hal self-selects and immediately provides an answer (wh-questions) to the teacher's (what is tested and how is tested) questions in the previous turn. However, his yet-in complete turn in line 5 is cut off by the teacher in a terminal overlap; therefore, it is unfortunately unavailable. In what follows, the teacher signals her continuation with a preface (so) with turn initial rising intonation and draws the attention to the distractors in the form of wh- questions (how (.)when (.)which (.) how often) by reading out loud. She marks the end of her turn with the confirmation token (right), which immediately receives an acknowledgment from Hal and Ela in sotto voice, respectively. After a gap of silence, this time teacher continues her turn in line 9 by reading the question stem (does he go to his work), yet she stops and employs an elongated change of state token (o:h) which displays her surprise on reading aloud the wrong question stem. She immediately issues a self-initiated self-repair with an open class repair initiator (sorry) and non-lexical perturbation (uhm) and reads the question stem (does he do his homework-at eight o'clock) in a rush-through. In the following turn, Hal establishes a mutual gaze with the teacher and prefaces his upcoming turn with the turn-initial possibility marker (maybe) and signals his upcoming suggestion to replace the problematic grammatical structure how often with another form of wh- question as the grammar section of the exam aims to test wh-questions and present simple tense in multiple-choice format. His offer is delivered as an example (maybe how often can be replaced with where) and marked with hedges maybe and (can), which downgrades his suggestion for the problem in the test item; therefore, it makes the next relevant action a rejection or acceptance of his suggestion as an SPP. With a pitch-rise in the last syllable of the candidate's suggested item (where) Hal fixes her gaze at the teacher, and with no time lapse, he receives an acknowledgment token (hmm hmm) embodied with a head nod (Gardner, 1995). After a gap of silence (0.2), he continues his turn-in-progress with the discourse marker (because) that indicates the start of an account-giving sequence. He further states that

the structure of ‘how often’ a distractor in the item should be replaced with where (because it is longer than other options and not the correct answer) because it is different in length from the other distractors and therefore violates a testing principle. While providing an immediate post-account (Waring, 2007) to his suggestion, he explicitly orients to the violated testing principle that “the length of the options should be kept about equal” (Haladyna & Rodriguez, 2013) and, at the same time, performs rule policing (Balaman, 2018). In the same turn, he adds with the discourse marker (so:) and orients to another principle to back up his claim (so it asks wh- questions) in the post suggestion position, which shows his orientation to the testing principle 'the point tested in the item is in line with the aim of the test (Fulcher, 2010; Heaton, 1990). Since the test intends to measure studenta’ ability on the grammar structure wh-questions and present simple interrogative. According to Can (2020), interactants' orientation to the testing principles during test-item reviewing interactions displays “their situated understanding of how a test should be and leads to its immediate acceptance” (Can, 2020, p. 492). from the others without further questioning. In line with his account based on the testing principles, the teacher delivers an acknowledgment token in elongated format (ohka:::y) in a terminal overlap and starts reading the distractors at a slower pace one by one and makes it available to the whole class. She stops before the trouble source in the item, which is the problematic grammatical structure ‘how often’. She increases her loudness with the elongated continuation marker (↑a:::nd), which receives a completion by Hal again with the repetition of his candidate suggestion as an insertion to the teacher's previous incomplete turn. In line 17, with the continuation marker (and), teachers employ an explicit positive assessment (Waring, 2018) to Hal's suggestion (this is a good suggestion) at the completion point of Hal's turn. Her positive assessment of the offered suggestion displays her agreement and affiliation with HAL. In the same turn, she further articulates the discourse marker (because) and provides another account for the reason to change ‘how often’. The teacher bases her account on the violated testing principle “overtesting” by explicitly stating the grammatical structure, how often, is already tested in the previous question with a deictic marker (here, in question 6) and therefore, it's not a suitable distractor in the question anymore. While she undertakes the problematization for the item from another perspective, HAS and BAR from the test-item group start taking notes, respectively. Their embodied

nodding and uptake for the confirmed change in the test item only starts after the teacher shows affiliation with HAL's offer, which displays their orientation to the teacher as the epistemic authority in the class. In line 19, the teacher continues her explanation with the discourse marker (now) and signals a transition to the suggestion phase in the form of 'obligation'' (now you have to find something else) after she reveals the problem in the item in line 18. Moreover, she addresses the test-writer group with the pronoun (you) and employs an if-conditional clause (if you have whe:re), which overlaps with Rey's turn initiation in line 21 with the possibility marker (maybe).

As one of the group members, Rey offers to reformulate Hal's suggestion by making all the options in the multiple-choice test item with the forms of -wh questions and taking out the questions beginning with 'how' from the distractors. Her reformulation of the announced change shows her orientation to the focal item as a member of the item-writer group who has the responsibility and right to ensure correct understanding to fix the problems accounted for in the test. This is also evident in her and other members' note-taking process during the feedback interaction. Her inclusive language (we) refers to her group members or the whole class. However, in line 23, the teacher only responds with the token (yeah) and does not show any orientation to assess Rey's suggestion. Her misalignment to this suggestion might arise from her acceptance of a change already suggested by Hal as one of the peer feedback groups. She continues with the contrastive marker (but) and again reformulates her problematization of the structure 'how often' and refers to the violated testing principle 'over-testing' on questions six and seven between lines 23-30. While she is providing an additional account to the violated principle, over-testing, which has been already problematized previously in session, she checks understanding with the confirmation token (right), which receives an immediate minimal response token (okay) as a claim of understanding from both Rey and Has. In line 32, the teacher signals the transition with the marker (now) to question 8 and finalizes the suggestion phase of the previously problematized item.

Extract 7 depicts how a previously problematized item in a multiple-choice test has been improved with the suggestions and collaborative negotiation of feedback

provider groups and the teacher in the first week of the peer feedback session. To address the problem in the construction of distractors, preservice teachers referenced the violated testing principles. Further, looked into the formulation of suggestions to offer a solution from the perspective of other testing principles that must be fulfilled. Different from extract 4, it has been clearly shown that preservice teachers also benefit from testing principles to support their claims while suggesting a change in the item. In this extract, the over-tested and structurally different option 'how often' is replaced with another wh-question 'where,' which is similar to other distractors in form and content. Therefore, the extract above indicates how preservice teachers employ 'testing principles' that they have added to their assessment literacy in the previous teaching sessions of the course to problematize and provides suggestions based on solid grounds during their feedback interaction. On the other hand, the teacher's alignment and agreement with the feedback provider groups indicate that their contribution is welcomed by the teacher and their peers, who continuously engaged in the note-taking process during the session.

4.3. Emergence of RTP in Resistance & Resistance Management Phase

In the upcoming extracts, the item-writer groups show resistance to the problematizations initiated by feedback groups.

Extract 8a.Group 4 /writing a recipe

1 T: **ohhkay then let's discuss that (2.1) now#**
#--1--->line 2
1:ELA raises her hand

2 T: **yes #hmm hmm>**
#--->

3 ELA: **er ı ı WONDER T:HAT e:r how they can use words(0.2)**

4 **in this picture for example er even if ı am an adult young**

5 **^adult ı ı can't ı cannot er:<u:se words like er:(.)φhamur>**
^--2---> φ--3--->

2: EDA shakes her hands and head

3:ela looks at teacher and shake her hands

6 T: **hm hmφ [DOUGH**

φ--->

7 ELA: **[ODOUGH YES DOUGH and OALSO:: [er**

O--4---> --->

4: Ela points to teacher

8 HAL: **e:r [its**

9 ELA: **[how can they Ωwhich kind of words they need to use whil-**

Ω--5---> line 18

5:T looks at HAL and points to him

10 HAL: **[its-cooking**

11 T: **[but Ωunintelligible unit**

12 HAL: **[+is the unit=**

+--6--- line 25

6:teacher nods her had

13 T: **=[hmm hmm they have cover[ed**

14 GUL: **[Yeah**

15 HAL: **[yeah they have taught the vocabulary=**

16 BUS: **=bir de şeyler var hani Ωblend (.) mix (.)**

there are also you know

Ω--->

17 EDA: **<Øelemek fe:lan (0.5) onları Øbiliyorlar mı**

do they know sifting and stuff like that

Ø--7--- Ø--->

7:Eda turns to Bus and makes sifting movement with her right hand

18 BUS: **ele:mek hepsi**

19 HAL: **hepsi yani [elemek felan var yani**

everything I mean there is sifting and so forth you know

20 BUS: **[χelemek yani bunlar var zaten**

sifting you know these already exist

--χ---> line 28

8:Ela turns and face BUS

21 HAL: **chopping var şey var [hepsi var yani**
there is chopping you know everything is available

22 BUS: **[°zaten hamur yapması yapması ingredientleri karıştırması**
besides making making dought mixing the ingredients and

23 **felan(0.3)**
so forth

24 **karıştırsın işte ing˜redientleri alsın mesela ya da vocabı**
let s/he mix well take the ingredients for example or

25 **eklesin de**
add the vocabulary

26 HAL: **YE:S +vocabı var**
the vocabulary is there
--->+

27 **ELA:[£helal χolsun ben kullanamıyorum buları£] (3.1)**
well done I cannot use these words
χ--->

Extract 8a begins with teachers signaling the transition to the discussion of group 4's writing section of the exam, which aims to test students' ability to write a recipe according to the picture prompts as nonlinguistic clues for the students. At the beginning of the interaction, there is a (2.1) second gap which displays the teacher's wait time for preservice teachers to show participation. Upon delay, the teacher remarks on her transition to the next discussion with the marker (now), and without further delay, Ela raises her hand and bids for a turn (Sandström, 2002) which receives an immediate acknowledgment for the next turn (yes hmm hmm) from the teacher. In the following line, Ela initiates her turn with the preface (I wonder) which signals her upcoming problematization with the writing task, recipe. She formulates the base sequence of her problematization with an ungrammatical wh-interrogative to challenge (Kooshik, 2002) the design of the writing task and invokes non-present actors (Leyland, 2021; Yöney, 2021) students, with the third personal plural pronoun (they). Ela directs her question concerning the ability of students to use words given in the pictures (how they can use words(0.2) in this picture.) While

formulating her problematization, Ela hints at a testing principle “difficulty level” (Can, 2020) while invoking non-present actors to take the matter from students' perspective. She continues her problematization and claims inability (even if I am an adult, I can't use words like er:) to use the expected target vocabulary in the writing task by invoking herself as a potential test-taker while nonverbally shaking her head (Mondada, 2007) in coordination with the modal verb (I can't I cannot). In the meantime, the teacher produces a continuer token (hmm hmm) and sustains her gaze at Ela when she initiates a word search through self-initiated repair and requests help for the word (hamur) while nonverbally shaking her hand repetitively, which displays her difficulty in eliciting the target word expected in the writing task. Her request for help for the trouble source and the word hamur act as complementary evidence to her claim that as an adult, it is difficult to use the expected words provided in the picture prompts of the question. Without any delay, the teacher provides the target word in English (dough) and receives an embodied confirmation token (yes) and repetition of the word (dough dough) with a rising intonation from Ela. In a terminal overlap, Ela signals her continuation with the discourse marker (also). At the same time, HAL, from item-writer group 4, produces an elongated hesitation marker (er:it's) which signals his readiness to take the next turn before mutually contacting eye gaze with the teacher. In line 9, Ela continues her yet-incomplete turn and directs another ungrammatical wh-interrogative (which kind of words they need to use while) at a fast pace, but her utterance cuts of by Hal and teacher in overlap to claim the next turn. As her problematization, FPP makes agreement or disagreement a second relevant action, and Hal and teacher both claim the next turn to respond. Between lines 10-12, the teacher and Hal chorally co-produce (Lerner, 1996a) the contrast marker (but), which signals an upcoming resistance to Ela's claim in the previous turn. In their choral co-production, Hal and the teacher express that the unit is about cooking (but cooking is the unit) from which the writing task is taken. As preservice teachers prepare their English language exams according to the books provided by the ministry of education, they must limit the content of their exams according to the topics and structures covered in the language books for each section they create. Therefore, in their choral co-production, Hal and the teacher frame their counterargument from teaching perspectives and books the students use in class. The teacher proceeds with her turn and acknowledges (hmm hmm)

Hal in an embodied way, and she brings further account by stating that the topic has been covered; therefore, the test measure what is taught in class (they have covered), which further displays her alignment and agreement with Hal's stance. In the next lines, as the item-writer group, Hal and his group member BUS start an extended multi-telling to provide counterarguments to Ela's problematization of the expected vocabulary in the writing task between lines 15 and 26. In this respect, Hal takes over to justify his claim, and he orients to the testing principle' test what you teach' by explicitly stating that students know the target vocabulary because they are taught (they have taught the vocabulary) in class in the cooking unit of the book. His statement implies that items tested in the exam must be taken from what students have covered in class; hence, they are responsible for what they are taught. Therefore, the writing task, recipe, is in the students' knowledge domain, and they can carry out the task as required. In her account-giving turn, Hal invokes non-present actors, students with the third person plural (they), as potential test-takers for the language test while orienting to the testing principle, which is fulfilled as students are assumed to be taught the target vocabulary in line with the teaching agenda of the language books. Hal's formulation of his account from the perspective of testing principles as a resource to be used while resisting the comment provided by one of the peer feedback groups indicates that preservice teachers display their knowledge and expertise in test item construction in different phases of the feedback interaction such as problematization, suggestion, and resistance.

In line 16, Bus shows affiliation to Hal's account-giving by exemplifying the target vocabulary that the students are assumed to know from the cooking unit in the book (bir de şeyler var hani blend mix). Upon Bus's statement, Ela turns her gaze and bodily orients to Bus to pursue further acceptance of her problematization and rushes to challenge her by directing another wh- question. This time, she asks whether the students know other target words like 'sifting' (elemek felan onları biliyorlar mı). Without any delay, Bus employs a full modified confirmatory repeat (elemek) of Ela's final utterance and further upgrades it with the indefinite pronoun (hepsi/all of them) to state that all the expected vocabulary is covered in the book. In line 19, Hal shows his agreement and alignment by producing a full modified repeat of BUS's previous turn (elemek felan hepsi). In line 20, Bus uses

the token (zaten/you know), which assumes knowledge as shared among the interactants, and repeats her exact statement in line 17 to reiterate her claim that the vocabularies presented in the pictures are all covered in the coursebook, and none of them is missed. In line 21, Hal employs a confirmatory repeat to Bus's statement, including another word (chopping) and clarifying that it is also another word provided in the coursebook. It has been indicated that after Ela's questioning about the target vocabulary expected from the students in line 3, Bus and Hal cooperatively worked on providing more examples as a counterargument to Ela's question.

Moreover, to justify their counterargument while orienting to the exam paper in front of them to provide the examples in line 22, Bus extends her statement and starts providing examples to the expected sentences that students are required to construct with the target vocabulary, such as (ingredientları karıştırması, hamur yapması). In line 26, Hal shows affiliation and agreement to his group member with the agreement token (yes) and a partial modified repeat of her final utterance (vocabı var). Following a moment of silence, Ela withdraws her gaze and turns her back to the item-writer group who is sitting behind her and signals her TCU termination with the idiomatic expression (helal olsun/well done), which is usually employed in topic closings in resistance interactions (Drew, 2005). However, her final TCU still restates her claim on the grounds that the vocabularies are difficult to use; therefore, it is not appropriate for the level the exam intended for. Moreover, she invokes herself again as a potential test-taker to justify the claim she formulated on her inability to carry out the task (ben kullanamıyorum bunları/I cannot use these). Therefore, her final TCU does not show her agreement with item-writer group members nor pursue further acceptance for her initial claim, as evident in her bodily behavior when she turns her head and stops contacting mutual eye gaze. Besides, she does not receive a response from the item-writer group nor the teacher as no one claims the turn-in (3.1) second-long gap. It is also noteworthy to include that neither her group members nor the teacher showed alignment to her claim raised on the difficulty level of chosen vocabulary for the writing task during the interaction. Thus, their lack of support displays that group agreement is absent in her attempt to problematize the writing task. In addition, the teacher shows nonverbal alignment to Hal and Bus as she nods during their extended telling to produce counterarguments and bring the account

to their resistance. Moreover, she was one of the first ones, alongside the group members, to show resistance to the problematization made in line 3.

It is discussed in 8a that Hal's orientation to the testing principle 'test what you taught' concerning the task' writing a recipe' functions as 'showing resistance to the claims raised on the task's design and difficulty level. Employment of testing principles' test what you teach' as an interactional source during a different phase of the feedback interactions shows that preservice teachers share a common ground in displaying fulfillment to the principles while constructing language tests. In showing resistance to the claims raised by their peers, item-writer groups provide counterclaims through the accounts to justify their design of the questions. Starting from line 12, Hal and Bus from the item-writer groups design their claims based on the testing principle, 'test what you teach and provide examples from the vocabularies taught in the class on a continuum that goes from general `(vocabı var)` to specific `(chopping, elemek, mix)`. As partners in the design of the item, they collaboratively work on an extended telling to show resistance and bring target vocabulary `(chopping, mix)` and expected sentence structure `(mix the ingredients)` that have been taught from the perspective of teaching and testing that creates a bridge between the learning and testing.

The extract displays that item-writers resort to RTPs for account-giving; however, they cannot bring an immediate resolution to the ongoing dispute between the two parties. This might be due to the fact that preservice teachers' epistemic status is equal regarding the employment of RTPs, which are shared epistemic sources in the class. Therefore, they are also negotiable objects (Firth, 1995). Moreover, pre-service teachers still will be evaluated on their language exams at the end of the course. Therefore, there might face issues and academic anxiety stemming from peer feedback, which is a new assessment approach for them.

At the end of the interaction, the preservice teacher who made the initial problematization did not withdraw her comment despite the resistance, which was supported and acknowledged by the teacher. Yet, the item is kept unchanged, which

indicates that the test is designed according to the testing principles and item-writer groups follow these rules while constructing language tests.

In the upcoming extract 8b, another preservice teacher from the peer feedback group initiates a problematization to the same writing task from another perspective: evaluation and answer key. The extract displays that preservice teachers' do not only provide general feedback to the items, but they evaluate the items from different perspective which improves the tests' quality and validity. In this extract, the item group 4's resistance does not receive an alignment from the teacher and fails at producing a valid counter-argument. At the same time, the teacher invokes larger pedagogical principles to address the new issue on the focal question as a learning opportunity for the whole class.

Extract 8b. Group 4/ Recipe

((selin raises her hand))

1 T: **SELININ? SORUSU VA::R hmm hmm**
sel has a question

2 SEL: **°you may write such an example er (0.2)[°in class]**

3 T: **↑[IN CLASS]**

4 SEL: **[ama mesela fotoğrafa göre eşleşmeyen cümle yazdığında**
but for example when they don't write matching answers

5 **da onu nasıl değerlendireceksiniz**
how are you going to evaluate

6 T: **HUH aynı benim sorum (0.5) teşekkürler**
it's exactly my question thanks

7 HAL: **°yani belli oranda esnetilebilmesine müsaade edicez °**
well we are gonna let bending of the rules to an extent

8 **tabii ki çünkü=**
of course because

9 T: **=HUH NASIL #nasıl nasıl yazıcaksınız#**
how how how are you going to write down

#teacher walks to HAL and stops#

10 HAL: **↑belli oranda esnetilmesine müsaade edebiliriz tabii ki**

we let them bend the rules at a certain extent of course

11 **yani revised key olduğu için**
I mean because we have a revised key

12 T: **hmm hmm**

13 BUS: **bütün öğrencilerin yazdıklarını oku[yara-**
by reading all students' writings

14 ELA: **[hepsine >put ve mix dese napıcaksın< yoksa er: aktivite**
what are you going to do if s/he say put and mix to each

15 **yani creative writing için hani↑ [olabilecek bi şey**
on the other hand the activity can be very well for creative writing

16 HAL: **[şimdi aslında (.) rubricte: bazı eksiklikler (.)var**
there are some missing points in the rubric

17 T: **[BEN DE olamaz mesela (0.2) aynı şekilde ne diyorlar diye**
I as well checked it for example what do they say

18 **baktım çünkü ↑mutlaka writingte bir tane EXAMPLE ANSWER (0.2)**
because for sure an example answer in the writing

19 **yani sizin <beklediğiniz ve sizin yazdığınız örnek bir şey>**
that you expect I mean an example that you expect

20 **+mutlaka yazılması lazım ideal answer +diye**
must be written as an ideal answer

+----------sibel nods head----------+

21 HAL: **[°hmm o şekilde°**
hmm in that way

22 T: **[baktım ama sizde yok**
I have checked it but you don't have it

23 SEL: **[°olması lazım bence°**
it should be I think

24 T: **[SADECE KEY VAR**
there is only a key

25 HAL: **sade:ce rubric üzerinden °gitmiş=**
it is only over the rubric=

26 T: **=EVET oraları sonra tartışırız**
YES we discuss those parts later

Prior to the extract, SEL raises her hand and bids for a turn (Sahlström, 2002) in the feedback session of group 4's writing task, recipe. In line 1, the teacher announces that SEL has a question for the item writer group with a rising intonation to draw the attention and allocates the turn to SEL through her acknowledgment token (hmm hmm). The item-writer group, Bus, Hal, and Gul, respectively, show bodily orientation to Sel and sustain their gaze before she starts her comment in line 2. In the following turn, Sel states that the recipe writing task might be practiced and taught in class; however, evaluation of the writing task, which is a well-known problem for the writing skill in English language tests, might still pose a problem for the item-writer group. During her formulation, Sel prefaces her statement by acknowledging the group's initial counter-argument regarding the "topic and vocabulary being taught" (you may write an example of the recipe in class), which the teacher emphasizes with the rising intonation through a partial modified repeat of her last utterance (in class). Her acknowledgment of the item-writer group's argument in the previous extract may act as a means to lay the ground for an upcoming new challenge instead of delivering it abruptly, which would probably result in an immediate resistance as in the previous extract. Without a time-lapse, SEL continues her turn with the contrastive marker (but) to signal her upcoming opposition (Can, 2020) and issues a wh- interrogative through invoking non-present actors, with the third person plural (they) to ask what will they do if the students provide unmatching sentences with the given pictures.

In line 6, the teacher shows agreement to Sel's challenge through her articulation of two compliance tokens respectively (huh exactly/aynen). In the same turn, she further thanks Sel for raising the issue and adds that she had the same question in her mind, which displays her agreement and affiliation with the comment and indicates that Sel's comment is a preferred contribution to the discussion. In the following turn, Hal starts her TCU with the uncertainty marker (yani/well) and states that they can let bending the rules for unmatching answers. But, his turn-in-progress cuts off by the teacher's repair initiation with the open class initiator (nasıl nasıl /how how), which signals her trouble in hearing as she starts walking toward Hal. In line 10, Hal

takes up his turn again with rising intonation and states that they can deal with the students''unmatching answers up to some point because they have a revised key in their hands. During his turn in progress, he receives an embodied acknowledgment token (hmm hmm) from the teacher.

Moreover, his group member Bus shows her alignment by adding that they will read and grade students' answers according to the revised key. Since preservice teachers are instructed to construct keys for every section in their exams, they come to the class prepared with their answer keys. Therefore, they bring the revised key as a justification to their account-giving to the claim initiated by SEL to problematize the marking of the writing task under review. However, Ela, who was the first to make a negative evaluation for the writing task earlier in the discussion, self-selects herself in line 14 again and raises another issue regarding the grading of the answers in addition to Sel's question in line 4. From her question, it can be inferred that she found the responses provided by Hal and Bus, respectively, insufficient considering the grading of the student's answers. While formulating her question by invoking non-present actors through the third person plural (they)to refer to students, she produces another question to challenge the item writer group regarding the key. She asks what they will do if the students give answers (Leyland, 2016) using a limited set of verbs such as mix and put. Through her enactment of students as potential test takers, Ela creates a hypothetical condition similar to Sel in line 4 to challenge the item writers' task design. She further adds that the writing task is an example of creative writing activity which provides a positive assessment (Pomerantz, 1984) to the item-writer group that might mitigate her negative assertion in the previous turn. Without further delay, Hal responds with the turn-initial (aslında/actually) and directly accepts that there are some missing points in their revised key. However, he does not show any orientation to the question raised by Ela, which displays that Ela was right in her raising the issue about the grading of the student's answers according to the revised key. Hal's acceptance of the problem regarding the incompleteness of the revised key further receives an additional comment from the teacher in the following turn. The teacher states that she also checked the design of the revised key and continues her turn by stating that writing an example answer to use an ideal writing model in the key is necessary for teachers. The teacher's comment treats the absence of example answers

in the revised key as a problem. In line 21, Hal claims understanding through the acknowledgment token (hmm o şekilde). Prior to teachers' advice-giving TCU between the lines 17-20, we can say that Sel's and Ela's questions to challenge the item-writer group resulted in the realization and acceptance of the additional problems in the test item concerning the marking of student answers.

After her statement is packed as advice, the teacher problematizes the lack of ideal answers as a model in their revised key corresponding to the writing task, recipe, in case students provide alternative sentences to the picture prompts. Her comment receives a partial modified repeat from Sel through the stance marker (bence/I think) with the last utterance of her turn (olması lazım). In line 25, Hal accepts the problematization and shows agreement with the teacher's comment by saying that the revised key is only based on the rubric. It is noteworthy that Hal's statement and acceptance of the teacher's comment regarding the key are formulated within the passive voice structure (rubric üzerinden gitmiş). While his formulation decreases his agency on the design of the task, it might be someone else from the group who prepared the key, and therefore Hal prefers to diminish his agency on the revised key. The teacher states that they will discuss the matter with the answer key later by following the order they follow in feedback-giving.

Extract 8b displays that the main focus in the discussion of a creative writing task, writing a recipe, has shifted from (test what you teach-extract 8a) onto the marking of the exam according to the design of the revised key. By challenging the item-writer group by formulating hypothetical scenarios regarding the possible student behavior (Yöney, 2021) feedback provider groups present additional problems that might arise due to the absence of an answer key which violates the principle “a criteria must be set for each unique writing task” (Heaton, 1990) to deal with the subjective marking problem of productive skills. It is seen in the analysis that, through the peer feedback groups' act of challenge, preservice teachers engage in negotiation and a continuous decision-making process to promote the validation of test items by pointing to the possible problems that might be encountered in the evaluation phase of the writing skill. According to Fulcher and Davidson (2007), “open discussion, negotiation, and

disagreement" during the review of test items is a key procedure for the validation process of the test items (Can, 2020).

The upcoming extract 8c follows the teacher's advice which receives a partial resistance from the item-writer group on its appropriacy to the students' level. The extracts indicates that even the advice is issued from the teacher, the preservice teachers still orient to every suggestion with a critical eye to evaluate its appropriateness on their language tests.

```
27  T:   (0.2) ama er: if IF YOU ARE creating such >exercises such
28       writing exercises< you should have remembered on the ielts
29       exam %they give an example answers on the toefl exam%
               %--8--->
                                                              %--->
    8:BUS NOD HER HEAD%
30  T:   they give example answers and you have to have especially
31       for the writing example answers all the time OKA.:Y?
32  GUL:→but hocam it's like çeight eight eight graders felan [SO::
                                ç--9---> line 34
         9:T walks towards GUL
33  HAL: [eight grade=
34  T:   =eight whatç
               ç--->
35  GUL: =eight #grades# so: may:be giving an example↓Ω
                                                        Ω--->
36  T:   NO NO you should not give the example to the students you as
37       teacher should have an example again to switch <to compare
38       the students answers> what do you really expect >this is why
39       ı am telling you that ↑>whenever you prepare a writing ex-
40       exercise first you sh- should SIT DO:WN and you should answer
41       the questions< (0.5)OHH:KAY? TH:EN you start creating this i-
42       this is the order that we follow (0.2) you create the exam(.)
43       you create the writing exercise (.) you sit down (.) create the
```

44 **answer(.)and ONLY THE:N depending on the answer(.)you create**
45 **the rubric (0.2) because if you ask me:: we have a huge**
46 **problem with the rubric here as well (0.2) because you expect**
47 **things that usually do not expect the rest**
48 HAL: **°okay°**
49 T: **O:KA:Y**

Following a micro pause (0.2), the teacher proceeds her turn with the contrast marker (but) and starts formulating an extended telling with the if-conditional structure (if you are creating such writing exercises) to directly address the item-writer group with the second person singular pronoun (you). In her turn in progress, she refers to the international language exams such as IELTS and TOEFL to support her advice in providing example answers, as they are also provided in famous international exams. During her multi-unit TCU, she also produces a reference to a past learning event (Can Daşkın, 2017; Can Daşkın & Hatipoğlu, 2019; You, 2015) on the international language exams by treating the information as studied in class before (you should have remembered). In the first week of the summer school, they covered the topic "different types of exams," where they checked national and international exams; therefore, the teacher treats the knowledge learned in her reference as a past shared event. At the termination of her TCU, she produces an understanding check through the token (okay) with a rising intonation. However, one of the preservice teachers from the item-writer group, Gul, signals her upcoming opposition to the advice provided in the previous turn with the contrastive marker (but). In line 32, Gul, another group member who didn't participate in the discussion earlier yet showed her engagement through her gaze sustained at the teacher, addresses the teacher (hocam/miss) to take her attention. Gul reminds the teacher that (it's eighth graders) by referring to the level stated on the exam specifications for the intended group of students the exam is designed for. In the following turn, Hal produces a full modified repeat of her turn (eighth grade). At the same time, the teacher employs an other-initiated repair through her repetition of the troubled word + wh interrogative (eight what) to signal her problem in hearing. Upon this, Gul repeats her initial comment with a louder intonation and states that giving an example for the eighth graders may not be appropriate. Besides, she prefaces

her turn through the hedge (maybe), which epistemically downgrades her claim. In her comment, Gul invokes the testing principle concerning "the item should be at a level appropriate to the proficiency level of test-takers" (Heaton, 1990). By orienting to the difficulty principle, she bases her opposition to the advice provided by the teacher on the reasonable ground regarding the difficulty level of the item, which she thinks violated the new suggestion delivered by the teacher (Can, 2020; Waring, 2017)

Yet, without any time lapse, the teacher issues the repair initiator component (no, no, no) at a fast pace and starts her multi-unit turn between lines 36-47 to deliver her advice. Yet, she states that her advice was to have an example answer on the key for teachers, not the students. Teachers' explanation of Gul's resistance displays that Gul misunderstood teachers' initial advice in line 17 about adding example answers. This time, the teacher elaborates on her advice by going general (Waring, 2017) with the adverb (whenever) and generic pronoun (you) and includes the whole classroom (we teachers) as a community in the delivery of the feedback. According to Waring (2017), through depersonalizing their advice by going general when carrying out the delicate work of providing feedback, teachers present the problem and, accordingly, the advice not to the specific person or group being criticized but to the whole community of people (we teachers) to secure agreement and alignment on the matter and manage any possible resistance.

Besides, the teacher delivers her advice step by step for the preparation of the writing tasks, by firstly mentioning the creation of the writing task for the exam, in the second turn answering the questions, and thirdly, rubric and the answer key for the evaluation.

In her final TCU, she restates that the item writer group has a big problem with the rubric, which receives a claim of understanding (Koole, 2010) through the acknowledgment token (okay) by Hal before the teacher issues an understanding check (okay) that come later in the interaction. It is worth noting that the teacher packs her advice as information and invokes a larger disciplinary principle after the item-writer group shows trouble understanding her first advice in line 31 (Waring, 2017). Preservice teachers nod their heads which cannot clearly indicate uptake of the advice

or not. Therefore, by going general she includes everyone in the classroom to create a learning space for them (Waring, 2017).

Extract 8c is the last analysis of the writing task, recipe, feedback session of group 4, which initially showed resistance to writing a recipe task by referencing the testing principle 'test what you teach' `(they have taught)`. Through their orientation to the testing principles, preservice teachers showed their understanding and expertise in writing English language exams, fulfilling the required principles. During their employment of testing principles, it is evident in their behavior that preservice teachers base their claims on accounts 'to establish a form of basis from which organizationally relevant action may be identified, challenged and discussed' (Firth, 1995). On the other hand, extract 8b displayed the emergence of other problems in relation to the evaluation and marking of the writing task under investigation through the negative evaluations made by other feedback group members. Preservice teachers' collaborative task of challenging the item-writer group through their questions led to the identification of other possible problems, such as marking alternative student answers, which is a well-known problem in evaluating productive skills such as writing and speaking, which are challenged with the 'subjective marking.' In this respect, preservice teachers' use of wh-questions `(what will you do, how will you evaluate)` constructed with hypothetical cases(`if they only write…)` to challenge (Kooshik, 2002) the item-writer group also resulted in teachers' extended- advice giving which also received resistance from one of the group members on account of 'difficulty principle' which she thought to be violated with the suggestion teacher offered for the revised key. However, not every account based on a principle is accepted directly by the group members. It is evident in the teacher's formulation of her advice by going general that accounts are also subject to negotiation and challenge (Firth, 1995), which is another way of ensuring mutual understanding on the matter.

The last extract in this section indicates the emergence of RTP in another sequential environment where peer feedback groups deal with the resistance initiated by item-writer groups.

Extract 9 is taken from the grammar session of the summer school. In the extract below, group 1, HAS, REY, BAR, and OZG receive feedback on their grammar session from group 4 HAL, BUS, ZEK, GUL, and group 5 MUS, MEH, AYS, NUR, SIB for the first question in their grammar section. The first question aims to test the simple present tense and is designed as a multiple-choice question for 6th-grade students.

Extract 9.Group 1 / elimination

1	T:	**%first &question%&**
		%looks at classroom%
		&ZEY raises his hand&
2	ZEY:	**it has a clue.**
3	T:	**hmm hmm**
4	ZEY:	**Ali likes riding bicycle but HE: (.)**
5		**of course er its an -s=**
6	T:	**ha ha:**
7	ZEY:	**we can er estimate it because by- by looking a-at the**
8		**first question**
9	T:	**o::ka:y?**
10	ZEY:	**°its not a clear (.) question (.)because of that°**
11	T:	**↑goo::d (.) so >you say that we have too many< likes (.)**
12		**in the questio:n**
13	ZEY:	**yeah also but (.) Ali likes (.) has the: answer of the**
14		**question (.) [actually**
15	T:	**[but he doesnt?**
16	HAS:	**[unintelligible]**
17	REY:	**but if you are unintelligible [students-**
18	ZEY:	**[you-you can get (.) you-you can get the**
19		**the s be-because of the Ali then you can ee:r delete**
20		**some eer options(.)by looking at that this looking=**
21	REY:	**=if you know simple present of course °will be better°**

22 **if you do not you can (0.2) you can just select Σd also**

(0.2) Σ--1---

>22

1:HAL looks at the exam paper in front of him

23 **ZEY: its just simple Σpresent tense=**

Σ--->

24 **REY: =but if the students ~d-don't~ know the students don't**

25 **know they [can't (0.2)**

26 **ZEY: [they should kno:w↑]**

27 **REY: [of course they should know but**

28 **SS: [£uh huh£]**

29 **REY: [e:r bişey sorabilir miyim]**

may I ask something

30 **S2: [unintelligible]**

31 **REY: [outline says that the only topic is ee:r simple**

32 **present tense.]how long (.) do you (.) prefer(0.2)er to**

33 **ask °this question°? how do you: (0.2) how would you ask**

34 **this question?**

35 ZEY: **er I am just suggesting that Ali likes should be**

36 **removed because of(0.3) e:r its an clue=**

37 T: **=hmm hmm**

38 BUS: **you measured but (.) sentence you don't measure simple**

39 **present tense**

40 ZEY: **yeah**

41 BUS: **you just measured (.) but**

42 ZEY: **yeap this is unintelligible [positive or negative**

43 BUS: **[do the students do the students know the meaning of but**

44 REY: **[the-this is not- this is just not the meaning of but]**

45 T: **bu:t ↑ also after HE:: ↑ (.) eer (.) they also measured**

46 **simple present tense [anyway right↑]**

47 REY: **[yes it is]**

48 T: **wh:y ↑ (0.2)**

49 ZEY: **↑#you told us that e:r (0.2) we shouldn't eliminate**
#sustains his gaze at teacher--->55

50 **er the options by looking at the question right=↑**

51 T: **=correct**

52 ZEY: **%thi-this helps students to eliminate% [(...)**
%-----points to the exam paper-----%

53 T: **[hmm hmm hmm hmm I knew (.) eer that and we should**

54 **rephrase in different manner (.) good and specific for**

55 **you I was just listening to you for a m- for now**

56 ZEY: **okay**

57 T: **#oka:y**
#--->

At the beginning of the extract, the teacher focalizes the first question in the grammar section of group 1 and sustains her gaze on the classroom, where Zey from the peer feedback group 3 raises his hand and bids for a turn (Sahlström, 2002). In the next turn, he immediately receives an embodied acknowledgment from the teacher (hmm hmm) and produces a preface (it has a clue) to his upcoming problematization by referring to a violated testing principle; the item has a clue to another item or the correct answer in the test (Haladyna, 2004). The violation of the principle in the question stem provides a clue that makes it easier for students to find the correct answer among the distractors of multiple choice-question. He continues his turn by reading aloud the first part of the sentence (Ali likes riding a bicycle but HE) and stops at the blank space, which aims to test simple present tense negative form (does not). Zey produces a certainty marker (of course) and provides the answer (it's an s), where he receives an embodied acknowledgment token from the teacher (hmm hmm) in coordination with an embodied head nod. In the next turn, Zey proceeds to problematize the focal question with inclusive language (we), referring to his group members or the whole classroom. He problematizes the design of the item and the question stem (Ali likes), which makes the estimation of the correct answer to easy to come up. In the next turn, he receives another acknowledgment token from the teacher (okay) and produces a negative assessment by stating that the question is

unclear. In line 10, the teacher nods and issues an explicit positive assessment (Waring, 2008), indicating that Zey's comment was a preferred contribution. She later demonstrates her understanding by reformulating Zey's previous comment (so you say) regarding the question stem providing a clue, and she brings an account to these clues (we have too many likes in the question) mentioned at the beginning of Zey's turn in line 2. In the next turn, Zey acknowledges and signals his continuation (also) to specify the clue he mentioned in the first part of the question stem (ali likes), which the teacher completes (but he does not) in a terminal overlap at the final utterance of his turn. In line 16, Has overlaps with Rey in the turn-initial position and hence withdraws her utterance, where Rey continues with the contrastive marker (but) to signal an upcoming resistance to the problematization made by Zey. In line 17, Rey invokes non-present actors (students) to formulate her resistance; however, her utterance cuts off when she overlaps with Zey's upcoming turn in line 18. Zey elaborates on his initial claim about the problematic question design in which the first part of the question structured with the present simple tense directly provides the answer to the blank space, which also tests the present simple tense and, therefore, can be easily estimated. In line 21, Rey responds and states that the question can only be answered if the students know the simple present tense structure, and if they do not, other distractors, such as option d, can also be selected. In her formulation, she formulates her resistance from the student's perspective and invokes non-present actors (students) to describe a possible student behavior while taking the test. (Leyland, 2021; Yöney, 2021) to show that the item cannot be easily estimated, which was Zey's initial claim. In line 23, Zey responds by commenting that the question only tests simple present tense structure, which must be in the student's epistemic domain (they should know) since it is the only grammar structure covered in class according to the test specifications written by the item-writer group. Doing this, Zey receives a shared laughter token from his peers, and in the upcoming turn, Rey issues a proforma agreement by employing a full modified repeat of Zey's previous comment (of course, they should know) yet issues a contrastive marker (but) to further pursue acceptance for her resistance. What follows is a pre-pre with the question (may I ask something), which precedes a challenge formulated to Zey by asking how he would test the same structure. During her formulation of the question, Rey's turn includes multiple micro pauses, hesitation markers, and repair in the same turn. In the next line,

Zey responds by suggesting that the first part of the question (ali likes) should be removed because it is a clue to the correct answer and downgrades his claim with the hedge (just). In the next turn, he receives an embodied acknowledgment token from the teacher, and in line 37, Bus, sitting next to Zey, starts her turn by directly addressing Rey with the second person singular pronoun (you). At the same time, Rey and Has, another member of the item-writer group 3, turn and sustain gaze at Bus. In her turn, Bus shows her alignment with her group member Zey by bringing additional support for the initial problematization of the item from another perspective. She states that the question tests (but) structure, not the simple present tense as it intends to do. In the next turn, Zey confirms his group member and shows his agreement by commenting that the question asks positive or negative forms with the connector (but). In the following line 43, Bus takes up her turn again and requests information by asking if the students know the connector (but). Her question reinforces her initial comment in line 39 and challenges Zey as the item writer. In the next turn, Zey performs a disconfirming response and directly rejects Bus's initial claim by stating that the question not only tests whether the students know the meaning of (but) or not. In the next turn, the teacher intervenes and states that the item-writers also measured the same structure, simple present tense after the connector (but he), which receives an immediate confirmation from Rey (yes, it is). Yet, the teacher directs a (why) interrogative to group 3 and invites elaboration on the design of the question. However, upon a micro (0.3) delay, Zey, who is the first to problematize the item, sustains his gaze at the teacher and starts his multi-unit turn by referring to a past shared moment (you told us), which directly takes the teacher as the main source for his upcoming comment. While formulating his statement, he makes a reference to the violated testing principle (we should not eliminate the question by looking at the question). His initial reference to the teacher shows his learning the rule regarding the elimination of the distractors in the multiple-choice questions. While he gives a reference to the rule by supporting his claim with the teacher's epistemic authority, he indicates that the rule is adopted from the teacher probably in the lectures given in the previous weeks of the course.

Before his turn completion, he issues a confirmation check (right↑) with a rising intonation. In the next turn, the teacher shows her alignment by confirming, and Zey

points to the exam paper (Mondada, 2007) with the deictic reference (this). In doing this, he invokes non-present actors (students) and reformulates his claim about the violated testing principle that the question stem helps students to eliminate (this helps students to eliminate). It is important to note that, as in the previous extracts analyzed in the problematization phase of the interaction, Zey also supports his problematization of the focal item by providing an account from a violated testing principle he issued in the pre-position of his sequence. He backs up his initial claim, rejected multiple times by the item-writer, by taking the teacher as his reference point, which indicates that he assumes the teacher as the authority and source of the principle he brought to the discussion. Following this, the teacher shows her alignment by acknowledging Zey in an embodied manner because she is the one directly addressed this time. Hence, she holds the right to confirm Zey's claim (Heritage, 2013) and thereby accept or reject the test item's problematization, which resolves the ongoing resistance between the groups. Simultaneously, she shows bodily orientation to group 3 as the item-writer group to produce a suggestion regarding rephrasing the problematic question stem. Her alignment with Zey and her suggestion to the group shows that Zey and Bus were right in their claims, and the teachers' continuation with the suggestion phase of the interaction shows that she accepted the problematization, which receives no further rejection from the item-writer group. However, the teacher does not directly address group 3 during her suggestion formulation, yet she uses inclusive language (we), which depersonalizes her advice delivery (Waring, 2007). The teacher proceeds with her comment and brings an account to Zey that she was well aware of the problem in the test item, yet she remained silent to listen to the peer groups' arguments during the feedback session. We can say that since the pedagogical aim of the lesson is to promote peer feedback during the sessions held for language exams. Since it was the first week of the feedback session, the teacher tailored her teacher talk by leaving room for preservice teachers to show participation and facilitate learner contributions. In line 54, Zey issues a minimal response token, and the teacher closes the turn with the turn-terminal (okay).

Unlike the previous resistance extracts, extract nine shows that testing principles are brought to the discussion as an interactional resource to manage an ongoing resistance from the item-writer groups. In her initial claim to problematize the question first, Zey

invokes different testing principles which are violated in the construction of the question stem. However, she fails to establish a mutual agreement between the peer groups. This demonstrates that, even though preservice teachers use RTPs for providing accounts, 'not every explanation is equally acceptable' Verkuyten (2000). On the other hand, showing resistance leads to a more extended negotiation among the preservice teachers, which only gets resolved by invoking the violated testing principle with reference to the teacher as a shared epistemic source and authority (Heritage, 2013; Sidnell, 2015) in the classroom. Zey's bringing the teacher and the violated testing principle in the pre-account position to back up his claim receives agreement from the teacher, which leads to the resolution of the resistance and acceptance of the problem. According to Firth (1995), accounts can function as problem-solving devices in such environments. However, this extract also shows that testing principles are still negotiable among preservice teachers. Thereby, the acceptance of the problematization depends on the kind of evidence they bring to the interaction.

After the teacher's official acceptance of the problem, what follows is the suggestion phase to eradicate the problem in the test item, which is initiated by the teacher and leads to the next phase in the peer feedback discussion.

The following chapter will discuss the findings presented in this section and discuss its implications for the existing and future research.

CHAPTER 5

DISCUSSION

This chapter discusses the findings in the previous section of the study based on four main research questions. In the first section, the sequential environments of RTP and their functions are discussed. In the following, referred testing principles are presented with a comparison to previous studies in the literature. Besides, the effect of RTP on creating learning opportunities for preservice teachers to improve their assessment skills is described. Finally, the study's implications for L2 teacher education and implementation of peer feedback in higher education are also addressed.

In the line with the first research question of the study, the in-depth analysis of peer feedback interaction among preservice teachers showed that peer feedback interaction mainly consists of 3 phases as follows 1) problematizing the test item, 2) showing resistance to the problematization, 3) suggestion for the elimination of the identified problem. The first phase in the peer feedback interaction is about the preservice teachers' problematizing the test items, which violates the testing principles and creates potential problems in the validity and reliability of the language exams. The second phase focuses on the emergence of resistance between the feedback provider groups and item-writer groups during the defense of their language exams against the announced problematization by orientating to the testing principles as an interactional resource. The last phase of the interaction focuses on preservice teachers' reference to testing principles while suggesting a possible solution to the problematic test items to eliminate the identified issues and enhance the quality of language tests by presenting their claims from the perspective that testing principles must be fulfilled.

In a similar vein to institutional context, it is seen in this research that preservice classroom interaction adopting a flipped classroom model to integrate the theoretical

and practical sides of the language testing and evaluation also has recurring phases of exchange which are shaped within the frame of "pedagogical task" that is, peer feedback**.**

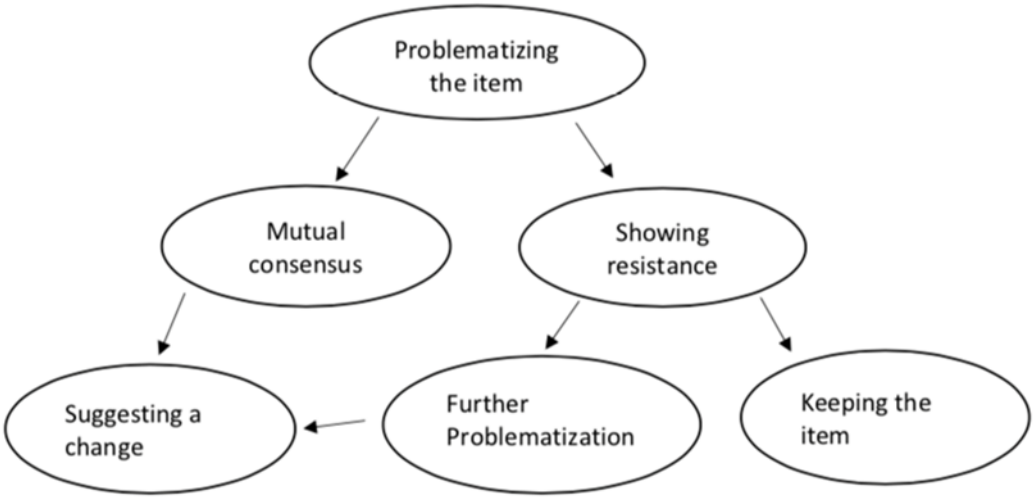

Figure 2 Sequential Environments RTP Emerges in the Peer Feedback Interaction

5.1. Emerging Recurrent Phases in the Peer Feedback Interaction

5.1.1. Problematization

The initial step in the peer feedback interaction is the problematization of test items through a continuous negotiation process beginning with the identification of the problematic item and account-giving by invoking a testing principle that is violated. The problematization phase leads to the recognition of possible problems and establishes mutual agreement on the issue as a threat to the validity and reliability of the language test.

As presented in the analysis section of the study, the phenomenon 'reference to the testing principle' mostly emerges in the peer feedback interaction during the problematization stage initiated by peer feedback groups. The analysis has shown that, among 77 instances of RTP, 38 examples are obtained from the problematization category. The frequent use of RTP as an interactional source during preservice teachers' evaluations of the focal test items might originate from the requirement to

back up their claims or assertions. According to Antaki and Leudar (1990), while presenting claims which are naturally disputable among the present parties, preservice teachers are in dire need of backup their claims on solid grounds to deal with the resistance or refutation that might arise. However, orienting to the testing principles is not just a mere task of backing up claims. Still, instead, it is a work undertaken by preservice teachers to bring up the common knowledge in testing and evaluation, which is available to preservice teachers from the previous lectures of the ELTE course. Therefore, while they refer to these principles, preservice teachers provide recognitional references to the source of these testing principles, as evident in their use of `(hocam you told us, Heaton says, Robert says that)`. This means that the epistemic source of these testing principles brought to the interaction mainly originates from external shared book sources and shared past instructional events (Can Daşkın, 2017; Can Daşkın & Hatipoğlu, 2019; You, 2015) on which preservice teachers have built their language assessment literacy. The fact that preservice teachers initiate a problematization of a test item by referring to these shared sources in their use of testing principles might display their "situated understanding of how a test should be" (Can, 2020). Their employment of testing principles to bring evidence to an announced problematization directs peer feedback groups to notice the arising issues and develop a mutual understanding of the nature of the problem. That is, orientation to the testing principles during the review of language test items might help preservice teachers more than the use of a ready-made checklist which can only answer to a particular type of test and hence neglect the context-specific needs of language tests. Therefore, the use of RTP during the critical evaluation of the test items might contribute to the detection of potential problems and improve the validation process of test items which paves the way for enhancing preservice teachers' learning by equipping them with necessary skills such as item-writing and item-reviewing as a critical part of their future profession.

To discuss the sequential environments where referencing to the testing principles emerges during the problematization of the test item, the analysis showed that the process starts with the initiation of the problematization. In the first phase, a peer feedback group comments on the focal question to identify a problem in the design of the question. It may receive an acknowledgment from the teacher during its

formulation process. In what follows, preservice teachers bring an account to the problematization by referencing the testing principles through which they express the nature and severity of the problem by indicating the violation of the rules in the same TCU. After the problematization is announced with RTP following in a pre or post-account position, the teacher is seen to be agreeing and aligning with the announced problem in her follow-up turns in which she also invokes "larger pedagogical principles" (Waring, 2017). In her follow-up turns, the teacher also reminds the breached rules of testing and presents the problems in the test items by "pronouncing them as not isolated or idiosyncratic but violating some fundamental understanding of the profession" (Waring, 2017). While the teacher's follow-up turn establishes mutual agreement on recognizing the problem, item-writer or peer feedback groups display their agreement with minimal or nonminimal response tokens.

All things said, the general layout of the peer feedback interaction could be summarized in figure 1 presented below. However, it is essential to note that when preservice teachers bring recognitional references to their account-giving process through testing principles, they place the violated testing principle in pre-account positions before they initiate the problematization sequence. Therefore, the sequential position of the phenomenon, RTP shows variation when preservice teachers bring shared epistemic sources (books) or authorities (teacher, authors) into the interaction.

I. Problematization of the item & Post account→ follow-up turn by the teacher/peer feedback groups → Mutual agreement → Suggestion

II. Pre-account with reference & Problematization of the item → Follow up turn by the teacher/peer feedback groups → Mutual agreement → Suggestion

Regarding the sequential unfolding of the RTP, the study shares similarities with previous research on the employment of 'rules.' To exemplify, the emergence and employment of RTPs as an interactional resource in peer feedback interaction is similar to Balaman (2016)'s study on preservice teachers. In his research, Balaman (2016) discovered preservice teachers' self-rule policing strategies regarding using the first language. While the students showed an orientation to this rule breach throughout

their online task interactions, they have also formulated new rules according to the needs of the ongoing online interaction. The intersection between the two studies can be that preservice teachers also orient to the rule breaches and employ a wide range of testing principles according to their needs in different sequential environments, as exemplified above.

5.1.2 Resistance

The announcement of the problematization in the test-item makes the second relevant action as either acceptance of the problem or rejection by the item-writer groups.

Thereby, the second phase in the peer feedback interaction is the resistance phase initiated by the item-writer groups to defend their exams in the face of an abrupt negative feedback provided by peer feedback groups which results in a longer negotiation and hence decision-making process for the focal test-item. To start with, 11 instances of resistance where item-writer groups resist the provided negative feedback with an orientation to the testing rules are obtained from the main collection of the study. However, not every resistance accepted as a valid argument in the face of the problematization. Therefore, while some of them provided valid objections to defend their exams some of them are rejected and dealt within the domain of testing principles. Therefore, the use of RTP in the resistance phase of the peer feedback interaction functioned as a problem-solving mechanism to bring a resolution to the resistance phase initiated by item writer groups and/or peer feedback groups. In this sense our study differs from the study conducted by Waring (2017). In her study Waring studied the advice-giving strategies of mentors by invoking larger disciplinary principles. She has discovered that invoking larger principles helps ensuring the mutual understanding between the parties. Moreover, larger disciplinary principles helped the mentor receive alignment from the teachers without causing tension and disagreement between the mentor and the teacher. In our study, invoking the testing principles function as “problem solving devices” (emphasis mine) in arising disputes. By orienting the principles, preservice teachers justify their claims and bring a resolution to the ongoing disagreements.

The layout of the resistance phase in the interaction starts with the announcement of the problem in the focal item which includes a negative evaluation or suggestion sequence. After the announcement of the problem the item-writer groups respond in two different ways; 1) either showing their resistance with an orientation to the testing principle they think their language tests fulfill or 2) showing their resistance from the students' or their self- perspective which results in resistance-management sequence initiation. Therefore, I will examine the structure of resistance sequences with two different figures summarizing the general layout of the interaction below. In the first case where, preservice teachers oppose the feedback provided by the peer feedback groups with an orientation to the testing principles that they assume their language test does not violate on the contrary fulfills. If their argument based on a testing principle is valid, they also receive an agreement and alignment from the teacher herself as it is exemplified below.

III. Problematization of the item → Showing Resistance with the testing principle → Resistance Resolution→ Keeping the test-item unchanged

The analyzed extracts 8a, 8b, and 8c include an instance of resistance to the negative feedback from a peer feedback group and to a suggestion formulated by the teacher. It is seen in extract 8a that one of the peer feedback groups provided negative feedback to a specific writing task in the last week of the feedback session and formulates her comment from the perspective of students. However, in the second turn, item-writer group showed resistance by orienting to a testing principle that they think their test does not violate. It is important to note here that there is a reciprocal relationship between the announced problem and oriented testing principles.

That means, the problematization of the item was based on the difficulty level of the task for the 6th graders and hence the principle oriented to oppose the feedback was "test what you teach principle" which defended the appropriacy of the task to the students' level on the grounds of 'being taught' in class. In doing this, item-writer groups provided an account to their design and appropriacy of the tests that is in line with the principles of language testing and assessment. In examples like this where the item-writer group opposed the feedback with a RTP, the teacher also showed

agreement and alignment to their counter arguments which lead to rejection of the feedback and keeping the item unchanged.

Apart from the first figure explained above, there are also examples of (extract 9) resistance responded without a RTP and hence ineffective in terms of rejecting the problematization. In these kinds of examples, use of RTP emerges at the resolution point by the peer feedback groups to manage the ongoing resistance.

As in the first example, the resistance unfolds after the announcement of the problem in the focal test-item. In the following, item-writer group display resistance from different perspectives which fails at eradicating the announced problem and hence leads into a longer negotiation process in which preservice teachers cannot establish mutual agreement on the matter. It is shown in extract 9 that if the consensus is not reached, the peer feedback group issues a testing principle to defend their rightfulness which receives the alignment and agreement of the teacher and thereby brings a resolution to the ongoing disagreement.

IV. Problematization of the item → Showing resistance to the feedback → Counter opposition with the testing principle → Resistance Resolution: Recognition of the problem

As it is presented in the figure, referencing to the testing principles may also emerge in environments where preservice teachers cannot establish consensus on the matter. Thereby, while item-writer groups defend their language exams to counter challenge the problematization, the peer feedback group can also manage this resistance by employing a RTPs which brings a resolution to the ongoing disagreement. Suffice to say, RTP can also function as a resolution device in terms of showing resistance and managing the resistance between opposing parties.

As it is exemplified above, the problematization sequences initiated by peer feedback groups (extract 2,4b,5,8) or by the teacher's turn allocation (extract 1,3,4a,6,7,9) are presented with accounts formulated within testing principles which fortify their claims and lead to a mutual consensus on the test-item. In the following, as problematization

FPP makes an agreement/rejection relevant in the second turn. An agreement by the teacher often includes also invoking larger pedagogical principles in the immediate follow-up turn (extract 6a,6b,7,8a,8b) or as a post-expansion to recap the overall feedback (extract 1,2,3,4). On the other hand, the response to the problematization by the item writer and peer feedback groups display variation. As extracts (1, 2, 4) shows, the item-writer groups display minimal response token by nodding heads and taking notes of the announced problems or suggestion. At the same time, extracts (5, 7) indicate that preservice teachers acknowledge the problem and further reflect on their item-writing process, which provides a nonminimal response. On the other hand, peer feedback groups also display alignment and agreement with the announced problems by providing additional accounts, which proves that violation of a testing principle leads to the violation of additional testing principles in the same test item as extract (1, 2, 3) represents.

To summarize, the problematization sequence cements the understanding of the nature and severity of the problem while establishing mutual understanding within the interaction.

5.1.3. Suggesting a Change

In the last phase of the peer feedback interaction, preservice teachers suggest a change to eliminate a problem that is collaboratively recognized as a threat to the validity and reliability of the test item. The analysis dedicated to the suggestion phase showed that suggestion sequences are initiated by peer feedback groups `(maybe we can, I would)` as well as the teacher `(ne yapsın, başka bir şey lazım)`.In the phases started by the peer feedback groups, preservice teachers orient to the testing principles in two different sequential environments; 1) back up their candidate's suggestion 2) challenge a candidate suggestion.

To start with the structuring of suggestion sequences, in the first place, the mutual consensus on the problematized item makes the next relevant action a suggestion to bring a solution for the problematic item. Therefore, it is seen in extracts (5, 6a, 6b, 7) that preservice teachers formulate their suggestions by employing a RTP in different

sequential environments. Since references to the testing principles pave the way for the immediate acceptance of the comments (extract 1, 2, 3, 4,) preservice teachers support their suggestions to the item-writer groups by presenting evidence from the testing principles to be fulfilled with their suggestions in a pre- (extract 6b) or post-account (extract 7) positions. Extract 6b displayed that a preservice teacher, HAL employs a RTP in a pre-account position for his upcoming suggestion to show that his suggestion will fulfill an aforementioned principle (we give a guideline, that's guided like-writing guided and specific instructions in the question stem principle) therefore providing a valid and appropriate solution to the problematic item. On the other hand, extract seven showed that by employing a RTP preservice, teachers justify their suggestions and mention the violated problem, which will be eliminated. (we can replace how often with where because it is longer than other options-options should be similar in forms principle) with their suggestions. In what follows, the teacher and the item writers' negotiation of the offered change by inserting it in the focal item leads to its acceptance and the closure of the sequence. Nonetheless, it is crucial to say that sometimes item-writer groups also offer suggestions to the identified problems in their exams which shows their full engagement in the item-reviewing process. To conclude, the general layout of the suggestion phase explained above is presented as follows in figure 5.

V. Candidate suggestion & Pre or Post Account → Negotiation of the item → Acceptance of the suggestion →

The second sequential environment where suggestions are accompanied by RTP is the challenging phase of the offered suggestions. Extract (6a) shows that after the suggestion is announced to eradicate the emergent problem in the item, preservice teachers evaluate the candidate suggestions and orient to the testing principles to bring evidence to the refutability of the suggestion, which violates another testing principle and hence leading additional problems in the test-item (it would be too easy for 6th graders, what if they choose the ones they know, option vermiyorduk). Therefore, challenging the offered suggestion contributes to the realization of the additional problems that might emerge in the item and hence leads to the cancellation of the suggestion, which does not provide a reasonable solution that

is in line with the principles of testing. In this respect, the study shows resemblance to Duran's (2017) study on at a higher education EMI context. In the study, Duran has discovered students' orientations to rule-breaches which is use of L1 in the classroom in their repair actions. In the classroom, students frequently formulated their own rules according to the needs of the interaction and 'classroom cohort'. In this sense, she concluded that rules are 'learner-generated' and constantly employed in the face of a breach from the general norm to encourage use of L2. Similarly in our study, preservice teachers also orient to a wide range of testing principles while evaluating the suggestions which deviate from the norms of testing and evaluation according to the needs of the ongoing interaction.

To describe in detail, the general layout of the challenging suggestion in the last phase of the peer feedback interaction is presented in figure 6 below as follows:

VI. Candidate suggestion → Challenging the suggestion through testing principles → Withdrawal of the suggestion → Initiation of new suggestion phase

During their endeavor to deal with the problematic items in the language exams, preservice teachers orient to the different testing principles either fulfilled or violated to make sure that 1) the candidate changes fulfill testing principles and hence eradicate the ongoing problem in item 2) the new announced change does not violate any other testing principle and hence does not produce additional problems in item 3) mutual agreement is established among the preservice teachers through their employment of testing principles which shows their 'situated understanding of how a test should be (Can, 2020).

5.2. Functions Performed by RTP

In line with the fourth research question, we must mention the functions RTP performs in different sequential environments throughout the sessions. Firstly, in different sequential environments RTP unfolds, it is seen that there is a reciprocal relationship between the referenced testing principles that are violated and the problematization brought forward. The RTPs mentioned in the problematization sequences have

functioned as accounts which are further explained within the same TCU by peer feedback groups in two different ways. To exemplify, after employing a RTP, peer feedback groups explained the problem in the item in detail by referring to their test taking process. They illustrated a possible student behavior or reaction to the problematic question by taking the exam as a student (Can, 2020) and hence brought additional evidence to their problematization (ben ne yazacağımı bulamadım, bunu da diyebiliriz, ben bulamadım). This recurrent interactional pattern was also observed in the teacher's follow-up turn, through which she supported peer feedback groups' comments (şimdi ben in total mı yazayım yoksa ayrı mı yazayim). It is essential to state that both peer feedback groups and the teacher assumed themselves as the end-users of the language tests and brought a description of a possible student behavior from their perspective (Can, 2020).

Additionally, after employing RTP within their problematization sequences, peer feedback groups have also invoked non-present actors (Leyland, 2021; Yöney, 2021) to exemplify possible student behavior in their explanations of their negative evaluations during the suggestion phase (what if they write) also resistance phase (they are sixth graders so) in the interaction. In this case, the analysis showed that preservice teachers adopt many different intuitional roles during the feedback practice, such as student, teacher, peer feedback provider, and lastly, test maker role. Nonetheless, the roles undertaken by preservice teachers are not predetermined. On the contrary, these roles show variation according to the specific aim of each phase which brings different responsibilities to the interactants.

To summarize the functions of RTP in different sequential environments, in the most recurrent phase in the feedback sessions, RTPs performed an act of account-giving to bring additional evidence to the negative feedback initiated by peer feedback groups which leads to its recognition and thus the establishment of mutual agreement. In the second turn, RTPs functioned as challenges made by item-writer groups to produce a counterargument in defense of their language exams against the problematizations made by peer feedback groups. In the last phase, RTPs justified the offered suggestion to eradicate the identified problems in the test items. In the same phase, they are employed by peer feedback groups or the item writer to evaluate the appropriacy and

validity of the suggestions, resulting in acceptance or rejection. In this sense, the findings are similar to Can ‘s (2020) study on teacher’s item-reviewing practices. In her study, Can (2020) has discovered the use of testing principles in three recurrent phases as well; problematization, suggestion, reviewing the change. During teachers’ feedback practices, they have evaluated the items by drawing on their testing experience and invoked testing principles to bring an account to their claims. In our study, preservice teachers have also orient to the testing principles in a similar way, however one difference emerged at the resistance phase. In the study, preservice teachers have orient to different types of testing principles to resist the negative evaluation from peer feedback groups and defend their language tests. The analysis has showed that, RTP do not only provide accounts to claims but they also function as ‘problem-solving or resolution-devices’ in the face of disagreements or disputes. This difference might stem from preservice teacher’s academic anxiety about ‘grading’ Because at the end of the summer school they will be evaluated on their exams and hence they might take the negative feedback as a threat to their GPA and also their assessment competency.

All in all, the use of RTPs shows that preservice teachers ground their claims on concrete evidence and common ground in testing, which they have built their language assessment literacy and item-reviewing/ writing skills as an essential part of their future profession. According to Waring (2017), invoking larger pedagogical principles further facilitates “exploring principled understandings and socializing the teacher into important disciplinary and pedagogical conduct and conceptualization.”

5.3. Testing Principles Oriented in Three Recurrent Phases of Peer Feedback Interaction

This section discusses the oriented testing principles in a different types of language tests in peer feedback interaction according to the third research question of the study. The analysis chapter of the study showed that preservice teachers orient to an extensive range of testing principles which mainly focus on item-writing in different test types (multiple choice, matching, fill in the blanks), item analysis (difficulty level, content) and lastly item-formatting (instructions, options) in three different sequential environments.

To begin with, preservice teachers orient to item writing and formatting issues in the language test items mostly during the problematization phase of the peer feedback interaction. Some of the most violated testing principles that preservice teachers focused on are the (1) odd one out (extract 9) and (2) more than one correct answer '(extract 3). These principles have 21 instances among 77 cases in the data set. They mainly emerged in multiple choice question types, which are known to be easy to evaluate but difficult to construct (Heaton, 1990). In the following, (3) options should be similar in forms (extract 1,7) and (4) instructions should be grammatical, structured, and guided for students (extract 2,4,5,) and (5) language should be appropriate to the level of students (extract 3) principles emerged in the construction of question stems and formatting distractors. This indicates that preservice teachers evaluate the focal test items through a wide range of testing principles but also focus on the grammar-related issues while constructing test items that are appropriate to the level of target students.

This demonstrates that preservice teachers make use of their language expertise and rely on their testing experience when they critically evaluate test items that are found to be problematic and, therefore, in dire need of a possible solution. At this moment, the suggestion phase comes into the picture when the preservice teachers employ RTPs in the validation process of the test items.

The analysis of the study showed that principles mostly referred to during the problematization phase of the interaction are recurrently employed in the suggestion phase. However, this time preservice teachers made use of testing principles for justifying and/or challenging the offered suggestions. The preservice teachers did not only refer to the violated principles in the offered suggestion (6) "excessive verbiage or clues to the correct answer must be avoided-window dressing" (extract 6a) but also addressed other testing principles (7) "instructions should guide and lead students towards a specific type of task"(extract 6b) that their suggestions would fulfill and eliminate the emergent problem in the test-item. These findings are in line with the study conducted by Can (2020) in terms of displaying how preservice teachers show "their situated understanding of how a test should be" when they construct and review the items by staying loyal to rules established within the domain of testing and

evaluation (p. 493). On the other hand, preservice teachers' challenging the suggestions through account-giving from testing principles are also parallel with Waring's (2017) study in terms of invoking rules "to convey or cement understanding as a matter of fact impervious to questioning" in the emergence of resistance.

In the resistance phase of the interaction, the analysis showed that preservice teachers employ testing principles that are different from the previous phases.

To clarify, preservice teachers mostly centered on (7) difficulty level of test items (extract 8c) and (8) test what you teach (extract 8a, 8b) principles which are mainly concerned with the content of the coursebooks and proficiency level of the target group that is described on the test specifications preservice teachers have prepared. Therefore, preservice teachers defended their language tests and provided counter arguments to the problematizations by considering their specific testing and learning context as well as the target level students in their domain of knowledge as item-constructors (Heritage, 2013). Therefore, the employment of different testing principles in the resistance phase of the interaction showed that the construction of language tests depends on the unique and specific language teaching and testing context. This demonstrates that, counter arguments based on testing principles provide valid evidence and hence influence the decisions and implications offered to the test-items (Can, 2020). These findings align with the study conducted by (Can, 2020) in terms of displaying preservice teachers' orientations to test items by taking the specific learning and teaching context into consideration while constructing language tests that align with the principles of testing and evaluation. Moreover, this also indicates that preservice teachers show awareness of the rules that must be fulfilled during their item construction process and hence succeeds in the elimination of upcoming problematizations which are found to be baseless. Their "adherence to the testing principles" (Can, 2020) results in the shared understanding and mutual consensus on the matter.

Table 5 Oriented Testing Principles in Three Sequential Environments

	Odd one out/Elimination	Options should be similar in forms The length of the options should be kept about equal (Hatipoğlu, 2009; Heaton, 1990).
	Only one correct answer	Each item should have only one correct answer (Carr, 2011; Hatipoğlu, 2009; Heaton, 1990).
	Over testing	The same item, structure or lexical verb should not be tested more than once (Heaton, 1990).
Problematization and suggestion phases	Window dressing	Excessive verbiage or irrelevant clues should be avoided in the item (Haladyna, 2004; Hatipoğlu, 2009; Heaton, 1990).
	Sufficiency of context	There should be sufficient context to convey the intent of the questions to the students (Brown, 1996). A list of acceptable answers should be provided for productive response items (Brown,1996).
Resistance phase	Level of difficulty	The item should be at a level appropriate to the proficiency level of students (Heaton, 1990).
	Test what you teach	The item should represent a specific content area covered in class (Carr, 2011; Haladyna & Rodriguez, 2013).

Suffice it to say, preservice teachers' reference to testing principles during feedback delivery is shaped according to the needs of the ongoing interaction. Further, it demonstrates 1) how they establish mutual understanding on the matter and (2) how they solve arising problems on the items, and lastly, (3) how they deal with a rising dispute among each other in the face of relational matters.

5.4. Learning Opportunities RTP Offers throughout the Peer Feedback Sessions

In compliance with the fourth research question of the study, learning opportunities provided by RTPs are discussed in preservice teachers' self-feedback and teacher's follow-up turns.

5.4.1. Preservice Teachers' Display of Uptake in Follow-Up Turns

During the feedback sessions, preservice teachers display uptake of the feedback in two different ways 1) in their follow-up turns after the feedback delivery and 2) in their self-feedback turns before the feedback delivery. To begin with the first one, preservice teachers demonstrated uptake of the feedback after the teacher shows her agreement with the peer feedback groups. Nevertheless, extracts 1,2,3 have shown that, preservice teachers' employment of RTPs did not require the alignment of the teacher in some cases and directly managed to get acceptance from the other preservice teachers, which eliminated the teachers' follow-up turn and delayed it till the end of the session. In such circumstances, preservice teachers displayed their acknowledgment and alignment with the feedback by mostly nodding their heads and taking notes in the problematization phase. On the other hand, they demonstrated uptake of the suggestions to the problematic test items by offering reformulations (extract 6b) and understanding checks (extract 8c). Besides, it is seen in some cases that the item-writer groups offered suggestions to the problematic items in their exams (extract 6b), which indicates their acceptance of the problems and their engagement in the process of dealing with the emerging issues to enhance the quality of test-items. In the resistance phase (extract 9), it is also seen that item-writer groups recognize and accept the problem initiated by peer feedback groups after an RTP provide evidence to the problem and resolved the ongoing dispute. Their initial resistance to the negative feedback and later acknowledgment of the problem through RTP shows a change in their status.

Therefore, item-writer groups' note taking, acceptance of the problematization, and finally presenting possible solutions to the emerging problems all indicate a change in their status.

5.4.2. Preservice Teachers' Display of Uptake in Self-Feedback Turns

As a rule, established in the classroom each item-writer group has a chance to provide self-feedback to their exams at the beginning of feedback sessions.

During these self-feedback turns, it has been seen that item-writers also use RTP to indicate the rule breaches and hence problems they have noticed in their exams while explaining the changes that they will make to fix those problems.

As an example to this, extracts (4a,4b) described how a preservice teacher, TUĞ showed uptake of the feedback provided to him by the teacher and peer feedback group through a RTP `(instructions should come first)` in the first feedback session. In week 5, during the group's self-feedback turn, TUĞ has resorted to the same testing principle `(and instruction should come first right)` to pinpoint a problem in their exams which indicated his uptake of the feedback and policing it as a rule that must be fulfilled in the formatting of test-items. Tuğ's orientation to the rule might be a "self-policing" as Balaman (2016) and Duran (2017) has also termed. This self-policing indicates his uptake of the feedback and displays a change in his epistemic status from (K-) to (K+), which is also treated by the teacher 'as learning. According to Seedhouse (2010), any change in the behavior of learners indicate a possible micro-moment of learning that is observable. In this respect, Tuğ's self-policing led him to notice and fix the problems in the design of his test and prepare more efficient and valid language tests that is in line with the principles of testing. More importantly, his self-policing indicates that the focal testing principle is added to his language assessment literacy as a part of disciplinary knowledge.

5.4.3. Teacher's Invoking Larger Pedagogical Principles in Her Follow-Up Turns

The data analysis has showed that, initial reference to the testing principles by preservice teachers leads to the further negotiations on the matter in teacher's follow-up turns. The teacher generally invokes larger pedagogical principles (extract 1,3,9,6,7) at the end of the peer feedback sessions where she recaps all the feedback provided by preservice teachers, and further elaborates on the problem to ensure

mutual understanding. On the other hand, she also offers suggestions to the problems which are in line with the testing principles. This demonstrates that the employment of RTP as a common interactional source opens a space for learning opportunities by the mutual orientation of the teacher and the preservice teachers. The role of RTP in facilitating learner opportunities is also evident in preservice teacher's use of reference to past learning events (hocam/miss you told us, Heaton says, Robert says that). Through reference to past learning events (Can Daşkın, 2017; Can Daşkın & Hatipoğlu, 2019) preservice teachers take the teacher and the testing books as shared epistemic sources to bring concrete evidence to their references. As extracts, 1, 4a, 4b, 4c,9 have shown they display their knowledge by orenting to the testing principles (hocam you told us we shouldn't eliminate the options by looking at the questions) and they reference from available shared sources. We can say that these sources come from their assigned course readings (Heaton, 1990; Haladyna, 2004) or the teacher who also invokes pedagogical principles in lecturing and/or feedback sessions all the time.

In this sense, we can say that learning opportunities created in the interaction through RTP emerges at different phases of the interactions. Besides, preservice teachers' start using the testing principle with a fine-tuned language (elimine edebiliyoruz-odd one out , optionlar farklı olmaz-options should be similar in forms-, birden fazla doğru oluyor-more than one correct answer, çok kolay olur-difficulty principle,) during the suggestion and resistance phases of the interaction. Their use of a fine-tuned language in the later weeks of the sessions shows that 'RTP' adds to their language assessment repertoire and become a part of their language assessment literacy.

In some cases, the teacher orients to the principles also referred by peer feedback groups (extract 4,6) in her immediate follow-up turns which indicates her agreement with the comments of peer feedback groups. Her invoking principles, usually comes after the minimal response tokens produced by item-writer groups. As it is seen in the analysis of the study (extract 2,3), sometimes item-writer groups may provide weak uptake of the feedback through their minimal response tokens which is delivered late. Therefore, the teacher invokes larger pedagogical principles to ensure alignment from

the teachers by also going general to include everyone in the classroom (Schwab, 2011). Teacher's elaboration on the issue facilitates reflective talk from the preservice teachers (extract 1) which aborts the "negative attributions" from the item-writer groups by going general (Waring, 2017). In this sense, the findings of the study is similar to Waring' s (2017) work on advice-giving in mentor-teacher talk in terms of invoking rules for account-giving process during the feedback delivery. Yet, what is different from her study is the fact that in our context, preservice teachers and the professor together invoke the testing principles as an interactional source in different environments for different purposes. As the extracts 4a, 4b and 6a have shown, the teacher and the peer feedback groups can mutually orient to the problems in the test items by referring to testing principles which leads to the changes in the item-writer group's epistemic status and learning state (extract 5) as well (Seedhouse, 2008). Therefore, the study adds to the literature in peer feedback studies by displaying the effect it plays on equipping preservice teachers with item-writing and reviewing skills which is a neglected aspect in ELTE courses in Turkey (Büyükkarcı, 2016; Hatipoğlu, 2017). Through receiving and delivering feedback preservice teachers engaged in hands-on experience which prepares them for their future profession that requires assessment skills to construct tests and evaluate the outcomes.

5.5. Implications

5.5.1. Implications For L2 Teacher Education and ELTE Course

This study offers insights into a flipped model L2 teacher education classroom, which successfully integrated theory and practice in and through social interaction. Preservice teachers have completed the assigned course readings at home and come to the class prepared to hold discussions for the exams they have constructed. As the first implication, this interaction between preservice teachers and the course instructor provided a context-rich learning environment in the language testing and evaluation course. The peer feedback interaction provided students with hands-on experience through which they engaged in constructive discussions to provide timely and specific feedback on each other's performance. These multilogue interactions offered

reflections on the problematic test items and resulted in improvements on language tests.

In this sense, peer feedback interaction has demonstrated its potential for "peer learning" as it is evident in preservice teachers' uptake of the feedback and changes in their epistemic status, which led to improvements in the language test constructions. We can say that implementation of peer feedback in a teacher education context directed students for negotiating their weaknesses and strengths which positively reflected on their performance in constructing valid test-items. In line with its aim, peer feedback has contributed to the students' performance by providing them with constructive and timely feedback which is often neglected in teacher reflective practices (See Chapter 2).

The second implication of this study unfolds with the phenomenon of RTP. Preservice teachers' orientations to the testing principles in the first weeks of the sessions resulted in the co-construction of rules which created a guideline for preservice teachers throughout the ELTE course. In this sense, RTP provided preservice teachers with specific guidelines on the o identification of potential problems in their test items that would otherwise go unnoticed (Can, 2020). Besides, RTPs are revealed and shaped according to the needs of the interaction (Duran, 2017), which resulted in the improvement of language tests and enhancement of item writing and reviewing skills.

To summarize, the most important implication of this study is that preservice teachers' feedback interactions have provided them with a 'hands-on experience' (Köksal, 2014) in practicing item writing and reviewing, which is an essential part of their future job yet mostly a neglected aspect in the structuring of the ELTE courses. While the previous research brings up the famous problem that "theory often fails to inform practice" (Johnson, 1996, p. 766) in the design courses for teacher education programs (Şahin & Subaşı, 2019), this study puts forward a model which enables preservice teachers to engage in "dialogic, reflective practices to enhance their assessment skills and help them to undertake the role of *assessor*" (Yeşilçınar & Kartal, 2020). In this sense, the study contributes to the existing literature on teacher education by presenting a flipped-classroom model in language testing and evaluation courses.

Furthermore, the study offers to increase the course hours dedicated to the ELTE course in teacher education programs. Since the course is usually offered only for one semester (Şahin, 2019), most topics are not comprehensively covered. They, therefore, result in an insufficient level of assessment literacy for preservice teachers. Thus, having face-to-face classes provides preservice teachers with a context-rich environment in which the dialogic talk creates learning space and opportunities to develop a sound knowledge of LAL. In conclusion, offering more courses on L2 testing and evaluation course designed in a flipped-classroom model has the potential to provide more favorable outcomes for the attainment of "sound assessment" knowledge and skills compared to the traditionally offered ELTE courses in most of the teacher education programs. In this sense, this study is different from previous literature on teacher education because it brings evidence on how to structure an undergraduate must course to fully benefit preservice teachers with a comprehensive practice in L2 assessment and actively participate in learning by evaluating their peers' performances.

To suffice to say, the design of the ELTE course in flipped classroom model for preservice teachers to have dialogic reflections among each other is crucial in preparing them for their future job, which requires performing assessment-related tasks in more than half of the time allocated to teaching. Although there is a general and significant tendency to design undergraduate courses in most higher education contexts traditionally, meeting the changing needs of teachers is of the utmost importance for the quality of education in the 21st century.

CHAPTER 6

CONCLUSION

This chapter presents an overall summary of the study and its limitations and offers suggestions for further research.

This study has investigated preservice L2 teachers' classroom interaction in a language testing and evaluation course in the ELT program at a state university in Turkey. Drawing on the conversation analytic approach, 12 hours long classroom data examination led to the discovery of the phenomenon "reference to testing principles" (RTP). Based on the main research questions, the study has focused on the and main functions of RTP in different sequential environments during peer feedback sessions.

According to the study's findings, preservice teachers have employed various RTPs in different sequential environments. The main functions RTPs fulfill are described in the following phases 1) bringing an account to the problematization, 2) bringing a resolution to the ongoing resistance 3) justifying the suggestions. Through referencing testing principles, preservice teachers "showed their situated understanding of how a test should be" (Can, 2020). By creating a guideline from the testing principles based on shared epistemic sources, preservice teachers formulated their feedback practices. They established a mutual understanding of the potential problems unfolding in the language tests. Throughout the evaluation of test items, RTP mainly emerged in the problematization and suggestion phases of the interaction, which fueled constructive discussions and created a space for learning opportunities in the development of language assessment literacy.

The study has proven the necessity of creating an authentic classroom environment where preservice teachers can practice test-item construction and reviewing by

engaging in peer feedback interactions. As an answer to the famous question "how to test" (Brookfield, 2017), the study suggests the implementation of a flipped-classroom framework within the design of ELTE courses to provide preservice teachers with abundant opportunities to socially interact and develop their assessment skills and hence disciplinary content knowledge (Schulman, 1987). This study contributes to the existing literature in terms of developing an understanding of how preservice teachers improve their assessment skills and literacy through interactions, which is often a neglected area in language assessment and testing research as well as L2 teacher education.

6.1. Limitations

The study has investigated the focal phenomenon in the successive four weeks of peer feedback sessions. Yet, the study has its limitations in the scope of the sessions included in the analysis chapter. To clarify, the last week of the summer school, session 8 is dedicated to the presentations of the final versions of the language tests. In this sense, the further use of the phenomenon RTP, might be explored to understand its overall impact on the improvements made in the language tests items' last version. Notwithstanding these limitations, the study successfully presents a model for the implementation of peer feedback in a higher education context for facilitating L2 interaction and providing learning opportunities.

6.2. Suggestions for Further Research

The present study makes several noteworthy contributions to the teacher education research by portraying the circumstances which provide abundant learning opportunities and space for preservice teachers through the implementation of peer feedback. As previously stated, the existing research on language testing and evaluation and teacher education largely focus on the investigation of pre-service and in-service teachers' language assessment literacy. Besides, the studies largely draw on the perceptions, expectations, and beliefs of language teachers. However, the process teachers go through in their development of LAL and acquisition of assessment skills remain as a neglect area. In this sense, this study contributes to existing literature by looking into the social interactions of preservice teachers in language testing and

evaluation course which is a neglected aspect in teacher education studies. Thereby, it is recommended that further research be undertaken in the investigation of classroom learning practices of teachers in their authentic classroom environment to deeply understand their learning process teachers go through in their undertaking the role of assessor from being a novice teacher.

REFERENCES

Akbari, R. (2007). Reflections on reflection: A critical appraisal of reflective practices in L2 teacher education. *System*, *35*(2), 192-207. https://doi.org/10.1016/j.system.2006.12.008

Allwright, D. (2005). From teaching points to learning opportunities and beyond. *TESOL Quarterly, 39(1), 9–31.* https://doi.org/10.2307/3588450

Amir, A., & Musk, N. (2013). Language policing: micro-level language policy-in-process in the foreign language classroom. *Classroom Discourse*, *4*(2), 151-167. https://doi.org/10.1080/19463014.2013.783500

Anker-Hansen, J., & Maria Andrée, J. (2019) Using and rejecting peer feedback in the science classroom: A study of students' negotiations on how to use peer feedback when designing experiments. *Research in Science & Technological Education*, *37*(3), 346-365. https://doi.org/10.1080/02635143.2018.1557628

Andrews, S. (2003). 'Just like instant noodles': L2 teachers and their beliefs about grammar pedagogy. *Teachers and Teaching*, *9*(4), 351–375. https://doi.org/10.1080/1354060032000097253

Antaki, C., & Leudar, I. (1990). Claim-backing and other explanatory genres in talk. *Journal of Language and Social Psychology*, *9*(4), 279–292. https://doi.org/10.1177/0261927x9094004

Ashenafi, M. (2015). Peer-assessment in higher education – twenty-first century practices, challenges and the way forward. *Assessment & Evaluation in Higher Education*, *42*(2), 226-251. https://doi.org/10.1080/02602938.2015.1100711

Azarnoosh, M. (2013). Peer assessment in an EFL context: Attitudes and friendship bias. *Language Testing in Asia*, *3*, 1-10. https://doi.org/10.1186/2229-0443-3-11

Badem-Korkmaz, F., Ekin, S., & Balaman, U. (2022). Pre-service language teachers' resistance to teacher trainer advice on task design for video-mediated L2 interaction. *Classroom Discourse, 13*(2), 212-230, https://doi.org/10.1080/19463014.2021.2020144

Bailey, K., & Brown, J. D. (1999). Learning about language assessment: Dilemmas, decisions and directions & new ways of classroom assessment. *The Electronic Journal for English as a Second Language, 4*(1), 258-381. Retrieved from https://tesl-ej.org/wordpress/issues/volume4/ej14/ej14r1/

Bailey, R., & Garner, M. (2010). Is the feedback in higher education assessment worth the paper it is written on? Teachers' reflections on their practices. *Teaching in Higher Education, 15(2),* 187–198. https://doi.org/10.1080/13562511003620019

Balaman, U. (2016). *A conversational analytic study on the development of interactional competence in English in an online task-oriented environment.* [Unpublished doctoral dissertation]. Hacettepe University.

Balaman, U. (2019). Sequential organization on hinting in online task oriented L2 interaction. *Text & Talk, 39*(4), 511-534. https://doi.org/10.1515/text-2019-2038

Balaman, U., & Sert, O., (2017). The coordination of online L2 interaction and orientations to task interface for epistemic progression. *Journal of Pragmatics, 15*, 115-129. https://doi.org/10.1016/j.pragma.2017.01.015

Baştürkmen, H. (2012). Review of research into the correspondence between language teacher's stated beliefs and practices. *System, 40*(2), 282-295.

Batlle, J., & Seedhouse, P. (2020). Affiliation and negative assessments in peer observation feedback for foreign language teachers professional development. *Applied Linguistics Review,* https://doi.org/10.1515/applirev-2020-0001

Becker, H. P. (1995). *Howard becker on education*. Open University Press.

Bishop, J., & Verleger, M. (2012, October-3). *Feedback effects: Comparing the change resulting from peer and TA feedback to student solutions of model-*

eliciting activities [Paper presentation]. Fronties in Education Conference, Seattle, USA. https://ieeexplore.ieee.org/document/6462434

Bishop, J., & Verleger, M. (2013, June -23). *Testing the flipped classroom with model-eliciting activities and video lectures in a mid-level undergraduate engineering course* [Paper presentation]. ASEE Annual Conference & Exposition, Atlanta, USA. https://ieeexplore.ieee.org/document/6684807

Blanton, W. E., & Moorman, G. B. (1987). The effects of knowledge on the instructional behavior of classroom reading teachers. *American Educational Research Journal, 18*(3), 343-361. https://doi.org/10.2307/1162667

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what teachers think, know, believe, and do. *Language Teaching*, *36*(2), 81-109. https://doi.org/10.1017/S0261444803001903

Borg, S. (2006). *Teacher cognition and language education.* London: Continuum.

Brindley, G. (1998). Outcomes-based assessment and reporting in Language learning programmes: A review of the issues. *Language Testing*, *15*(1), 45–85. https://doi.org/10.1191/026553298668589981

Brookfield, S. D. (2017). *Becoming a critically reflective teacher*. Jossey-Bass.

Brookhart, M. S. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice, 8*(2), 153-169, https://doi.org/10.1080/09695940123775

Brookhart, M. S. (2010). *How to assess higher-order thinking skills in your classroom.* ASCD.

Brookhart, S. M. (2017). *How to give effective feedback to your students* (2nd ed.). ASCD.

Brown, H. D. (1996). Resources in language testing and assessment. *TESOL Quarterly, 30*(4), 781-781. https://doi.org/10.2307/3587936

Brown, J. D., & Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing*, *25*(3), 349–383. https://doi.org/10.1177/0265532208090157

Bulunuz, N., & Bulunuz, M. (2016). Öğretmen adaylarının mesleki gelişimi için iyi öğretmenlik uygulamaları: Klinik danışmanlık modeli. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi, 29*(2), 401-429. Retrieved from https://dergipark.org.tr/tr/pub/uefad/issue/26859/282485

Burns, A. (1996). Starting all over again: From teaching adults to teaching beginners. In D. Freeman, & J. C. Richards (Eds.), *Teacher learning in language teaching* (pp.154-177). Cambridge University Press.

Büyükahıska, D. (2020). İngilizce öğretmenliği adaylarının gözünden İngilizce ölçme ve değerlendirme dersi. *Rumelide Dil ve Edebiyat Araştırmaları Dergisi*, *5,* 581–598. https://doi.org/10.29000/rumelide.752539

Büyükkarcı, K. (2018). Assessment beliefs and practices of language teachers in primary education. *International Journal of Instruction, 7*(1), 107-120. Retrieved from https://dergipark.org.tr/tr/pub/eiji/issue/5136/69989

Calkins, L., Wiens, P. L., & Quinn, L. (2020). Examining student teacher reflection using video-based structured reflection procedures. *Educational Research*, *31*(1), 16-21. https://doi.org/10.20429/ger.2012.090101

Calkins, S., Grannan, S., & Siefken, J. (2019). Using peer-assisted reflection in math to foster critical thinking and communication skills. *PRIMUS, 30*(4), 475–499. https://doi.org/10.1080/10511970.2019.1608608

Can Daşkın, N. (2017b). *A conversation analytic study of reference to a past learning event in L2 classroom interaction: Implications for informal formative assessment.* [Unpublished doctoral dissertation]. Middle East Technical University.

Can Daşkın, N., & Hatipoğlu, Ç. (2019). Reference to a past learning event in teacher turns in an L2 instructional setting. *Journal of Pragmatics*, *142*, 16–30. https://doi.org/10.1016/j.pragma.2018.12.023

Can, H. (2020). *A micro-analytic investigation into EFL teachers' test item reviewing interactions* [Unpublished doctoral dissertation]. Middle East Technical University.

Cañabate, D., Nogué, L., Serra, T., & Colomer, J. (2019). Supportive peer feedback in tertiary education: Analysis of pre-service teachers' perceptions. *Education Sciences*, *9*(4), 280.-296. https://doi.org/10.3390/educsci9040280

Carson, J., & Nelson, G. (1996) Chinese students' perception of ESL peer response group interaction. *Journal of Second Language Writing*, *5,* 1-19. http://dx.doi.org/10.1016/S1060-3743(96)90012-0

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL Instructors classroom assessment practices: Purposes, methods and procedures. *Language Testing*, *21*, 360-389. https://doi.org/10.1191/0265532204lt288oa

Cizek, G. J., Fitzgerald, S. M., & Rachor, R. A. (1995). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, *3*(2), 159–179. https://doi.org/10.1207/s15326977ea0302_3

Clayman, S. E. (2002). Disagreements and third parties: Dilemmas of neutralism in panel news interviews. *Journal of Pragmatics*, *34,* 10-11. https://doi.org/10.1016/s0378-2166(02)00070-x

Clift, R. (2006). Indexing stance: Reported speech as an interactional evidential. *Journal of Sociolinguistics, 10*(5), 569-595. https://doi.org/10.1111/j.1467-9841.2006.00296.x

Crandall, J. A. (2000). Language teacher education. *Annual Review of Applied Linguistics*, *20*, 34–55. https://doi.org/10.1017/s0267190500200032

Crook, A., Mauchline, A., Maw, S., Lawson, C., Drinkwater, R., Lundqvist, K., Orsmond, P., Gomez, S., & Park, J. (2012). The use of video technology for providing feedback to students: Can it enhance the feedback experience for staff and students*? Computers & Education, 58*(1), 386–396. https://doi.org/ 10.1016/j.compedu.2011.08.025

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*, 438-481. https://doi.org/10.3102/0034654305800443

Cushing, A., Abbott, S., Lothian, D., Hall, A., & Westwood, O. (2011). Peer feedback as an aid to learning – What do we want? Feedback. When do we want it? Now!. *Medical Teacher*, *33*(2), 105-112. https://doi.org/10.3109/0142159x.2011.542522

Davies, A. (2013). Fifty years of language assessment. In K. Antony (Eds*.), The companion to language assessment* (pp.1-19). J. Wiley., & Sons Inc.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*(3), 393-415. https://doi.org/10.1002/j.1545-7249.2009.tb00242.x

Demirbulak, D. (2012). Fostering propenstities for reflective practice at undergraduate methodology, and practicum courses. *Euroasian Journal of Educational Research, 48*, 39-58.

Demirel, E., & Enginarlar, H. (2016). Effects of combined peer-teacher feedback on second language writing development. *Hacettepe University Journal of Education*, *31*(4), 657-675. https://doi.org/10.16986/huje.2016015701

deSouza, D. K., Emma, B., Clinkenbeard, M., Morita, E., Shrikant, N., & Tuccio, W. A. (2021). Taking a detour before answering the question: Turn-initial okay in second position in English interaction*. Language & Communication, 76,* 47–57. https://doi.org/ 10.1016/j.langcom.2020.09.005

Dinkelman, T. (2000). An inquiry into the development of critical reflection in secondary student teachers. *Teaching and Teacher Education*, *16*(2), 195–222. https://doi.org/10.1016/s0742-051x(99)00055-4

Doehler, P., & Lauzon, F. V. (2015). Documenting change across time: Longitudinal and cross-sectional CA studies of classroom interaction. In N. Markee (Ed.), *The handbook of classroom discourse and interaction* (pp. 425–439). Wiley Blackwell.

Doughlas, D. (2010). *Understanding language testing*. Routledge

Drew, P. (1991). Asymmetries of knowledge in conversational interactions. In I. G. Fulcher, & F. Davidson (Eds.), *Language Testing and Assessment* (3rd ed., pp.156-192). Springer.

Drew, P. (2005). Conversation analysis. In K. L. Fitch & R. E. Sanders (Eds.), *Handbook of language and social interaction, 8*(3), 71-102. Lawrance Erlbaum Associates Publishers.

Drew, P., & Heritage, J. (1992). Analyzing talk at work: An introduction. In P. Drew & J. Heritage (Eds.), *Talk at work: Interaction in institutional settings* (pp. 3-65). Cambridge University Press.

Duran, D. (2017). *Student-initiated questions in English as a medium of instruction classrooms in a Turkish higher education setting* [Unpublished doctoral dissertation]. Hacettepe University.

Duruk, E. (2016). Deeper insights into university students' perceptions about peer feedback in second language writing: A phenomenological approach. *Journal of Qualitative Research in Education, 4*(3), 28- 42. https://doi.org/10.14689/issn.2148-2624.1.4c3s2m

Ellis, V., Edwards, A., & Smagorinsky, P. (2010). *Cultural-historical perspectives on teacher education.* Routledge.

Engin, M. (2014). Macro-scaffolding: Contextual support for teacher learning. *Australian Journal of Teacher Education, 39*(5), 26-40. http://dx.doi.org/10.14221/ajte.2014v39n5.6

Escobar-Urmeneta, C. & Walsh, S. (2017). Classroom interactional competence in content and language integrated learning. In A. Llinares, & T. Morton (Eds.), *Applied linguistics perspectives on CLIL* (pp. 189- 206). John Benjamins.

Erdemir, N., & Yeşilçınar, S. (2021). Reflective practices in micro teaching from the perspective of preservice teachers: Teacher feedback, peer feedback and self-reflection. *Reflective Practice*, *22*(6), 766–781. https://doi.org/10.1080/14623943.2021.1968818

Erkmen, B. (2014). Novice EFL teachers' beliefs about teaching and learning, and their classroom practices. *Eğitim Fakültesi Dergisi*, *29*(1), 99-113.

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, *83*(1), 70–120. https://doi.org/10.3102/0034654312474350

Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in education and training international*, *32*(2), 175–187. https://doi.org/10.1080/1355800950320212

Falchikov, N. (2013). *Improving assessment through student involvement*. Taylor and Francis.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70*(3), 287–322. https://doi.org/10.3102/00346543070003287

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research, 38*, 47-65. https://doi.org/10.1080/0013188960380104

Farrell, T. S. C. (2012). Review of reflective writing for language teachers. *Electronic Journal of Foreign Language Teaching, 11*(2), 358- 359. Retrieved from http://e-flt.nus.edu.sg/

Farrell, T. S. C., & Ives, J. (2014). Exploring teacher beliefs and classroom practices through reflective practice: A case study. *Language Teaching Research*, *19*(5), 594–610. https://doi.org/10.1177/1362168814541722

Firth, A. (1995). Accounts in negotiation discourse: A single case analysis. *Journal of Pragmatics*, *23*(2), 199-226. https://doi.org/10.1016/0378-2166(94)00003-W

Fitch, K. L., & Sanders, R. E. (2014). *Handbook of language and social interaction*. Psychology Press, Taylor & Francis Group.

Freeman, D., & Johnson, K. E. (1998). Reconceptualizing the knowledge-base of Language teacher education. *TESOL Quarterly*, *32*(3), 397-415. https://doi.org/10.2307/3588114

Freeman, D., & Richards, J. C. (1993). Conceptions of teaching and the education of Second language teachers. *TESOL Quarterly*, *27*(2), 193-208. https://doi.org/10.2307/3587143

Frodden, M., Restrepo, M., & Maturana, L. (2004). Analysis of assessment instruments used in foreign language teaching. *Ikal a Revista de Lenguaje y Cultura, 9*(1), 171-201. Retrieved from https://revistas.udea.edu.co/index.php/ikala/article/view/3146

Frunza, V. (2014). Advantages and barriers of formative assessment in the teaching-learning activity. *Procedia-Social and Behavioral Sciences, 114*, 452-455. https://doi.org/10.1016/j.sbspro.2013.12.728

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

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, *9*(2), 113-132. https://doi.org/10.1080/15434303.2011.642041

Gardner, R. (2004). On delaying the answer: Question sequences extended after the question. In R, Gardner, & J. Wagner (Eds.), *Second language conversations* (pp. 126). Continuum Books.

Gardner. R. (2007). The right connections: Acknowledging epistemic progression in talk. *Language in Society*, *36*(3), 319–341. http://www.jstor.org/stable/4169609

Genç, E., Çalışkan, H., & Yüksel, A. P. (2020). Language assessment literacy level of EFL teachers: A focus on writing and speaking assessment knowledge of the teachers. *Sakarya University Journal of Education*, *10*(2), 274–291. https://doi.org/10.19126/suje.626156

Ghafarpor, H. (2016). Classroom conversation analysis and critical reflective practice: Self-evaluation of teacher talk framework in focus. *RECL Journal*, *48*(2), 1-16. https://doi 10.1177/0033688216631173

Giraldo, F. (2019). Language assessment practices and beliefs: Implications for language assessment literacy. *HOW Journal, 26*(1), 35-61. https://doi.org/10.19183/how.26.1.481

Giraldo, F., & Murcia, D. (2019). Language assessment literacy and the professional development of pre-service language teachers. *Colombian Applied Linguistics Journal*, *21*(2), 243-259. https://doi.org/10.14483/22487085.14514.

Gonen, K., & Akbarov, A. (2015). Instructors' principles and practices in the implementation stage of classroom-based language assessment in higher education in Turkey. *Journal of European Education*, *5*(3), 28-39. https://doi.org/10.18656/jee.04624

Göker, S. D. (2006). Impact of peer coaching on self-efficacy and instructional skills in TEFL teacher education. *System, 34*(2), 239-254. https://doi.org/10.1016/j.system.2005.12.002

Guardado, M., & Shi, L. (2007). ESL students' experiences of online peer feedback. *Computers and Composition, 24*, 443-461. https://doi.org/10.1016/j.compcom.2007.03.002

Gwyn–Paquette, C., & Tochon, F. V. (2002). The role of reflective conversations and feedback in helping preservice teachers learn to use cooperative activities in their second language classrooms. *The Modern Language Journal*, *86*(2), 204–226. https://doi.org/10.1111/1540-4781.00145

Haddington, P. (2006). The organization of gaze and assessment as resources for stance taking. *Text & Talk, 26*(3), 281-328. https://doi 10.1515/TEXT.2006,012

Haladyna, T. M. (2004). *Developing and validating multiple-choice items* (3rd ed). Lawrance Erlbaum Associates.

Haladyna, T. M., & Rodriguez, M.C. (2013). *Developing and validating test-items*. Routledge.

Hale, C. C., Nanni, A., & Hooper, D. (2018). Conversation analysis in language teacher education: An approach for reflection through action research. *Hacettepe University Journal of Education*, *33,* 54-71. https://doi.org/10.16986/huje.2018038796

Hall, G., & Danby, S. (2003, December-3). *Teachers and academics co-constructing the category of expert through meeting talk* [Paper presentation]. The

Australian Association of Educational Research in Education Annual Conference, Auckland, New Zealand.

Halliday M. A. K. (1967). *Intonation and grammar in British English.* Mouton.

Hanlon, N., Murray, D., & Niculae, E. (2020). Student peer feedback and assessment in higher education. *Practitioner Research Projects*. 27. Retrieved from https://arrow.tudublin.ie/ltcpgdprp/27

Hasselgreen, A., Carlsen, C., & Helness, H. (2004). *European survey of language testing and assessment needs. Part one: General findings*. http://www.ealta.eu.org/documents/resources/survey-report-pt1.pdf

Hatipoğlu, Ç. (2009). Writing and evaluating multiple choice items. A training module prepared for the E-INSET Project. http://e-inset.mersin.edu.tr/

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), 12,* 40-51.

Hatipoğlu, Ç. (2015). English language testing and evaluation (ELTE) training in Turkey: Expectations and needs of preservice English language teachers. *ELT Research Journal, 4*(2), 111-128. Retrieved from https://dergipark.org.tr/en/pub/eltrj/issue/28780/308006

Hatipoğlu, Ç. (2017), History of language teacher training and English language testing and evaluation (ELTE) education in Turkey. In Y. Bayyurt, & N. C. Sifakis (Eds.), *English Language Policies and Practices in the Mediterranean Countries and Beyond* (pp. 227-258). Peter Lang.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Heritage, J. (1984a). *Garfinkel and ethnomethodology*. Polity Press.

Heritage, J. (1984b). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 299-345). Cambridge University.

Heritage, J. (2002). Oh-prefaced responses to assessments: A method of modifying agreement/disagreement. In C. Ford, B. Fox, & S. Thompson (Eds.), *The language of turn and sequence* (pp. 196-224). Oxford University Press.

Heritage, J. (2012). Epistemics in action: Action formation and territories of knowledge. *Research on Language & Social Interaction*, *45*(1), 1–29. https://doi.org/10.1080/08351813.2012.646684

Heritage, J., & Raymond, G. (2005). The terms of agreement: Indexing epistemic authority and subordination in talk-in-interaction. *Social Psychology Quarterly, 68*(1), 15–38. https://doi.org/10.1177/019027250506800103

Heritage, J., & Roth, A. (1995). Grammar and institution: questions and questioning in the broadcast news interview. *Research on Language and Social Interaction, 28*(1), 1-60.

Heritage, M., & Heritage, J. (2013). Teacher questioning: The epicenter of instruction and assessment. *Applied Measurement in Education, 26*(3), 176-190. https://doi.org/10.1080/08957347.2013.793190

Hobbs, V. (2007). Faking it or hating it: Can reflective practice be forced? *Reflective Practice, 8*(3), 405-417. https://doi.org/10.1080/14623940701425063

Huisman, B., Admiraal, W., Pilli, O., van de Ven, M., & Saab, N. (2018). Peer assessment in MOOCs: The relationship between peer reviewers' ability and authors' essay performance. *British Journal of Educational Technology*, *49*(1), 101-110. https://doi.org/10.1111/bjet.12520

Huisman, B., Saab, N., Van Den Broek, P., & Van Driel, J. (2018). The impact of formative peer feedback on higher education students' academic writing: A Meta-Analysis. *Assessment & Evaluation in Higher Education, 44*(6), 863-880. https://doi.org/10.1080/02602938.2018.1545896

Huisman, B., Saab, N., Van Driel, J., & Van Den Broek, P. (2019). A questionnaire to assess students' beliefs about peer-feedback. *Innovations in Education and Teaching International*, *57*(3), 328–338. https://doi.org/10.1080/14703297.2019.1630294

Hyland, K., & Hyland, F. (2006). Feedback and second language students' writing. *Language Teaching, 39*(2), 83-101. https://doi.org/10.1017/S02614448060003399

Inbar-Lourie, O. (2012). Language assessment literacy. *The encyclopedia of applied linguistics*. https://doi.org/10.1002/9781405198431.wbeal0605

İnan Karagül, B., Yüksel, D., & Altay, M. (2017). Assessment and grading practices of ELF teachers in Turkey. *International Journal of Academy, 5*(18), 168-174. https://doi.org/10.18033/ijla.3656

İnce, B. (2016). *The perceptions of pre-service ELT teachers on different modes of peer feedback and its relation to teacher efficacy* [Unpublished doctoral dissertation]. Middle East Technical University.

Jacobs, G.M., & Farrel, S.C.T. (2001). Paradigm shift: Understanding and implementing change in second language education. *Teaching English as a Second or Foreign Language, 5*(1), 1-17.

J.B. Heaton. (1990*). Writing English language tests*. Longman.

Jefferson, G. (1978). Sequential aspects of storytelling in conversation. In J. Schenkein (Eds.), *Studies in the organization of conversational interaction* (pp. 219-248). Academic Press.

Jenkins, A., R, D., Watts, S., Clifford, N. J., Stainfield, J., Bradbeer, J., Chance, J., Healey, M., Foster, I., Livingstone, I., Matthews, H., Barrett, H., Browne, A., Riley, D., Foord, J., Lindsay, J., Wareing, J., Longworth, D., The Curriculum Committee, Boothby, J., & Jackson, S. (1994). Control and independence strayegies for large geography classes. *Journal of Geography in Higher Education*, *18*(2), 245–262. https://doi.org/10.1080/03098269408709264

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. https://doi.org/10.1177/0265532209351431

Johnson, K. E., & Freeman, D. (2001). Teacher learning in second language teacher education: A socially-situated perspective. *Revista Brasileira De Linguística Aplicada*, *1*(1), 53–69. https://doi.org/10.1590/s1984-63982001000100004

Johnson, K. E. (2009). *Second language teacher education: A sociocultural perspective*. Routledge.

Johnson, R. T. (1994). David Johnson: A leading teacher of cooperative learning. *Teaching Education*, *6*(2), 123–125. https://doi.org/10.1080/1047621940060215

Kääntä, L. (2012). Teachers' embodied allocations in instructional interaction. *Classroom Discourse*, *3*(2), 166–186. https://doi.org/10.1080/19463014.2012.7166

Kääntä, L., & Lehtinen, E. (2016). Patterns of experience talk and argumentation in digital peer learning discussions. *Language@Internet, 13*(2), 1-17.

Kagan, D. M. (1988). Teaching as a clinical problem solving: A critical examination of the analogy and its implications. *Review of Educational Research*, *58*(4), 482-505. https://doi.org/10.3102/00346543058004482

Kasper, G., & Wagner, J. (2011). A conversation-analytic approach to second language acquisition. In Atkinson, D. (Ed.), *Alternative approaches to second language acquisition* (pp. 117–142). Routledge.

Kayacan, A., & Razı, S. (2017). Digital self-review and anonymous peer feedback in Turkish high school EFL writing. *Journal of Language and Linguistic Studies, 13*(2), 561-577. Retrieved from https://dergipark.org.tr/tr/pub/jlls/issue/36120/405630

Kiesling, S. F. (2011, April-15). *Stance in context: Affect, alignment and investment in the analysis of stance taking* [Paper Presentation]. iMean conference 2nd Conference, Bristol, UK.

Kızılay, E. (2018). Fen bilgisi öğretmen adaylarının mühendislik kavramına ilişkin bilişsel yapılarının incelenmesi. *Journal of Social and Humanities Sciences Research*, *5*(27), 32-38.

Kim, H. (2002). The form and function of next-turn repetition in English Conversation**.** *Language Research, 38*(1), 51-81.

Kim, Y., & Silver, R. E. (2016). Provoking reflective thinking in post observation conversations. *Journal of Teacher Education*, *67*(3), 203–219. https://doi.org/10.1177/0022487116637120

Khalil, E. (2018). The efficacy of peer feedback in Turkish EFL students' writing performance. *Journal of Literature and Art Studies*, *8*(6), 920-931. https://doi.org/10.17265/2159-5836/2018.06.011

Kleinsasser, R. C. (2005). Transforming a post-graduate level of assessment course: A second language teacher educator's narrative. *Prospect, 20*(3), 77-102. https://doi.org/10.3316/aeipt.149126

Koç, B., & Ilya, A. (2016). Exploring pre-service language teachers' perceptions and actual practices of giving feedback in micro-teaching. *Procedia-Social and Behavioral Sciences*, *232,* 421-429. https://doi.org/0.1016/j.sbspro.2016.10.058

Koole, T. (2010). Displays of epistemic access: Student responses to teacher explanations. *Research on Language and Social Interaction, 43*(2), 183–209. https://doi.org/10.1080/08351811003737846

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*(4), 344 348. https://doi.org/10.1016/j.learninstruc.2009.08.005

Koshik, I. (2002). Designedly incomplete utterances: A pedagogical practice for eliciting knowledge displays in error correction sequences. *Research on Language & Social Interaction*, *35*(3), 277–309. https://doi.org/10.1207/s15327973rlsi3503_2

Köksal, D. (2004). Assessing teachers' testing skills in ELT and enhancing their professional development through distance learning on the net. *Turkish Online Journal of Distance Education (TOJDE), 5*(1).

Kuşçu, S. (2014). Türkçenin yabancı dil olarak öğretiminde öğretmen bilişi ve öğrenci inanışları. *Hacettepe Üniversitesi Yabancı Dil Olarak Türkçe Araştırmaları Dergisi,* (1), 111-120. Retrieved from https://dergipark.org.tr/tr/pub/huydotad/issue/37783/436219

Lazaraton, A., & Ishihara, N. (2005). Understanding second language teacher practice using microanalysis and self-reflection: A collaborative case study. *The Modern Language Journal*, *89*(4), 529–542. https://doi.org/10.1111/j.1540-4781.2005.00328.x

Lee, Y. (2006). Respecifying display questions: Interactional resources for language teaching. *TESOL Quarterly, 40*(4), 691–713. https://doi.org/10.2307/40264304

Lerner, G. H. (1996a). On the 'semi-permeable' character of grammatical units in conversation: Conditional entry into the turn space for of another speaker. In Ochs, E. A. Schegloff. & S. A. Thompson (Eds.), *Interaction and grammar*, (pp. 238-276). Cambridge University Press.

Lerner, G. H., & Kitzinger, C. (2007). Extraction and aggregation in the repair of individual and collective self-reference. *Discourse Studies, 9*(4), 526–557. https://doi.org/10.1177/1461445607079165

Leyland, C. (2016). 'Pre-enactment' in team-teacher planning talk. Demonstrating a possible future in the here-and-now. *Pragmatics, 26*(4), 675-704. https://doi.org/10.1075/prag.26.4.07ley

Leyland, C. (2018). Resistance as a resource for achieving consensus: adjusting advice following competency-based resistance in L2 writing tutorials at a British University. *Classroom Discourse, 9*(3), 267-287, https://doi.org/10.1080/19463014.2018.1480966

Leyland, C. (2021). The interactional construction of the academic reader in writing tutorials for international students: An advice-giving resource. *Linguistics and Education*, *61,* 1-13. https://doi.org/10.1016/j.linged.2020.100900

Liddicoat, A. J. (2007). *An introduction to conversation analysis.* Continuum.

Liston, D. P., & Zeichner, K. M. (1987). Critical pedagogy and teacher education. *Journal of Education, 169,* 117-137.

Li, L. (2020). *Language teacher cognition: A sociocultural perspective*. Palgrave Macmillan.

Li, L., & Walsh, S. (2011). 'Seeing is believing': Looking at EFL teachers' beliefs through classroom interaction. *Classroom Discourse, 2,* 39-57. http://dx.doi.org/10.1080/19463014.2011.562657

Liu, N., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher Education*, *11*(3), 279-290. https://doi.org/10.1080/13562510600680582

Llosa, L. (2011). Standards-based classroom assessments of English proficiency: A review of issues, current developments, and future directions for Research. *Language Testing*, *28*(3), 367–382. https://doi.org/10.1177/0265532211404188

Lynch, B. K., & Davidson, F. (1994). Criterion-referenced language test development: Linking curricula, teachers, and tests. *TESOL Quarterly, 28*(4), 727-743. https://doi.org/10.2307/3587557

Markee, N. (2000). *Conversation analysis (second language acquisition research series).* Routledge.

Markee, N. (2005). Conversation analysis for second language acquisition. *Handbook of Research in Second Language Teaching and Learning*, 379–398. https://doi.org/10.4324/9781410612700-30

Markee, N. (2008). *Conversation analysis*. Routledge.

Markee, N. (2019). *The handbook of classroom discourse and interaction (blackwell handbooks in linguistics)*. Wiley-Blackwell.

Markee, N., & Kasper, G. (2004). Classroom talks: An introduction. *The Modern Language Journal*, *88*(4), 491–500. https://doi.org/10.1111/j.0026-7902.2004.t01-14-.x

Markee, N., & Seo, S. (2009). Learning talk analysis. *IRAL - International Review of Applied Linguistics in Language Teaching, 47*(1), 37-63. doi:10.1515/iral.2009.003

McCarthy, J. (2017). Enhancing feedback in higher education: Students' attitudes towards online and in-class formative assessment feedback models. *Active*

Learning in Higher Education, *18*(2), 127–141. https://doi.org/10.1177/1469787417707615

Mendonça C.O. and Johnson, K.E. (1994) Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly*, *28,* 745-769. http://dx.doi.org/10.2307/3587558

Merrifield, M. R., & Saari, D. G. (2009). Telescope time without tears: a distributed approach to peer review. *Astronomoy and Geoghypsics*. *50*(4). 16-20. https://doi.org/10.1111/j.1468-4004.2009.50416.x

Mertler, C. A., & Campbell, C. (2006, April-11-15). *Measuring teachers' knowledge & application of classroom assessment concepts*: *Development of the assessment literacy inventory.* [Paper presentation]. The Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Metin, M. (2010). Biçimlendirici değerlendirmeye yönelik öğretmen adaylarının düşünceleri. *Milli Eğitim Dergisi, 40*(187), 293-310. Retrieved from https://dergipark.org.tr/tr/pub/milliegitim/issue/36197/407045

Mondada, L. (2018). Multiple temporalities of language and body in interaction. *Research on Language and Social Interaction, 1,* 85-106. https://doi.org/10.1080/08351813.2018.1413878

Morawski, M., & Budke, A. (2019). How digital and oral peer feedback improves high school students' written argumentation: A case study exploring the effectiveness of peer feedback in geography. *Education Sciences*, *9*(3), 178. http://dx.doi.org/10.3390/educsci9030178

Munoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in and ELF context in Colombia. *Profile, 14*(1), 143-158.

Nassaji, H. (2000). What's the use of 'triadic dialogue'?: An investigation of teacher-student interaction. *Applied Linguistics, 21*(3), 376–406. https://doi.org/10.1093/applin/21.3.376

Nelson, M., & Schunn, C. (2008). The nature of feedback: how different types of peer feedback affect writing performance. *Instructional Science*, *37*(4), 375-401. https://doi.org/10.1007/s11251-008-9053-x

Nguyen, H. T., & Ngo, N. T. (2017). Learning to reflect through peer mentoring in a TESOL practicum. *ELT Journal*, *72*(2), 187–198. https://doi.org/10.1093/elt/ccx053

Nguyen, L. T. (2016). *Development of Vietnamese pre-service EFL teachers' assessment literacy* [Unpublished doctoral dissertation]. Victoria University of Wellington.

Nicol, D., Thomson, A., & Breslin, C. (2013). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education*, *39*(1), 102-122. https://doi.org/10.1080/02602938.2013.795518

Nilson, L. (2003). Improving student peer feedback. *College Teaching*, *51*(1), 34-38. https://doi.org/10.1080/87567550309596408

Nilson, L. B. (2013). *Creating self-regulated learners: Strategies to strengthen students' self-awareness and learning skills.* Stylus Publishing.

O'Loughlin, K. (2006). Learning about second language assessment. Insights from a postgraduate student on-line subject forum. *University of Sydney Papers in TESOL*, *1,* 71-85.

O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, *30*(3), 363–380. https://doi.org/10.1177/0265532213480336

Öz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal, 6*(1), 25-44. Retrieved from https://dergipark.org.tr/tr/pub/eltrj/issue/28779/307996

Park, I. (2014). Stepwise advice negotiation in writing center peer tutoring. *Language and Education, 28*(4), 362-382, https://doi.org/10.1080/09500782.2013.873805

Park, Y. (2014). The roles of third-turn repeats in two L2 classroom interactional contexts. *Applied Linguistics*, *35*(2), 145–167, https://doi.org/10.1093/applin/amt006

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

Pekrun, R. (2005). Progress and open problems in educational emotional research. *Learning and Instruction, 15,* 497-506. https://doi.org/10.1016/j.learninstruc.2005.07.014

Perakyla, A. (1997). Conversation analysis: A new model of research in doctor-patient communication. *Journal of the Royal Society of Medicine, 90*(4), 205-208. http://dx.doi.org/10.1177/014107689709000406

Pomerantz, A. (1980). Telling my side: Limited access as a fishing device. *Sociological Inquiry, 50*(3-4), 186-198. https://doi.org/10.1111/j.1475-682X.1980.tb00020.x

Pomerantz, A. (1984). Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes. In J. M. Atkinson, & J. Heritage (Eds.), *In structures of social action: Studies in conversation analysis* (pp. 57–101). Cambridge University Press

Ratminingsih, N. M., Artini, L. P., & Padmadewi, N. N. (2017). Incorporating self and peer assessment in reflective teaching practices. *International Journal of Instruction*, *10*(4), 165–184. https://doi.org/10.12973/iji.2017.10410a

Raymond, G. (2003). Grammar and social organization: Yes/No interrogatives and the structure of responding. *American Sociological Review*, *68*(6), 939–967. https://doi.org/10.2307/1519752

Rodman, G. J. (2010). Facilitating the teaching-learning process through the reflective engagement of pre-service teachers. *Australian Journal of Teacher Education*, *35*(2), 20-34. http://dx.doi.org/10.14221/ajte.2010v35n2.2

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal*, *59,* 23-30. https://doi.org/10.1093/elt/cci003

Rowntree, D. (2015). *Assessing students: How shall we know them*? Routledge

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696–735.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119-144. http://dx.doi.org/10.1007/BF00117714

Sahlström, F. (2002). The interactional organization of hand raising in classroom interaction. *Journal of Classroom Interaction*, *37,* 47–57. Retrieved from http://www.jstor.org/stable/23870411

Saraç-Süzer, H. S. (2007). Öğretmen adaylarının 8. sınıf İngilizce ders kitabındaki diyalogların etkinliği konusunda inançları: Toplumedim sorunları. *Kastamonu Eğitim Dergisi, 15*(1), 401-410. Retrieved from https://dergipark.org.tr/tr/pub/kefdergi/issue/49108/626729

Sato, K., & Kleinsasser, R. C. (2004). Beliefs, practices, and interactions of teachers in a Japanese high school English department. *Teaching and Teacher Education*, *20*(8), 797–816. https://doi.org/10.1016/j.tate.2004.09.004

Schegloff, E. A. (2007). *Sequence organization in interaction*. Cambridge University Press.

Schiffrin, D. (1987). *Discourse markers.* Cambridge University Press.

Seedhouse, P. (2004a). *The interactional architecture of the language classroom: A conversation analysis perspective*. Blackwell.

Seedhouse, P. (2004b). Conversation analysis methodology. *Language Learning*, *54*(S1), 1–54. https://doi.org/10.1111/j.1467-9922.2004.00268.x

Seedhouse, P. (2005). Conversation analysis as research methodology. In K. Richards & P. Seedhouse (Eds.), *Applying Conversation Analysis,* (pp. 251-266). Palgrave.

Seehan, S., & Munro, S. (2017, July-6). *Assesment: attitudes, practices and needs. BAAL language learning and teaching* [Paper presentation]. BAAL Language Learning and Teaching SIG, Lancashire, Preston.

Serdar Tülüce, H. (2019). The correspondence between beliefs and practices of an EFL teacher regarding L2 grammar teaching. *Cumhuriyet International Journal of Education, 8*(3), 791-806.

Sert, O. (2015). *Social interaction and L2 classroom discourse.* Edinburgh University Press.

Sert, O. (2019). The Interplay between collaborative turn sequences and active listenership: Implications for the development of L2 interactional competence. In M. R. Salaberry & S. Kunitz (Eds.), *Teaching and testing L2 interactional competence: bridging theory and practice* (pp. 142-166). Routledge.

Sert, O., & Balaman, U. (2018). Orientations to negotiated language and task rules in online L2 interaction. *ReCALL, 30*(3), 355-374. https://doi.org/10.1017/S0958344017000325

Sert, O., & Seedhouse, P. (2011). Introduction: Conversation analysis in applied linguistics. *Novitas- ROYAL, 5*(1), 1-25. Retrieved from https://dergipark.org.tr/tr/pub/novroy/issue/10815/130376

Sevimel-Sahin, A., & Subasi, G. (2019). An overview of language assessment literacy research within English language education context. *Kuramsal Eğitimbilim Dergisi [Journal of Theretical Educational Science]*, *12*(4), 1340-1364.

Shaw, C., Potter, J., & Hepburn, A. (2015). Advice-implicative actions: Using interrogatives and assessments to deliver advice in mundane conversation. *Discourse Studies, 17(3), 317–342.* https://doi.org/10.1177/1461445615571199

Shepard, L. A. (2005) The role of assessment in a learning culture. *Educational Research*, *29*(7), 4-14. https://doi.org/10.12691/education-4-9-9

Sidnell, J., & Stivers, T. (2013). *The handbook of conversation analysis*. Blackwell

Sidnell, J., & Stivers, T. (2014). *The handbook of conversation analysis*. Wiley Blackwell.

Spiller, D. (2012, February 12). *Assessment matters: Self-assessment and peer assessment*. The University of Waikito. https://cei.hkust.edu.hk/files

Stiggins, R. J. (1991a). Assessment literacy for 21th century. *The Phi Delta Kappan*, *77*(3), 238-245.

Stiggins, R.J. (1991b). Assessment literacy. *The Phi Delta Kappan*, *72*(7), 543-539.

Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice, 18,* 23-28. https://doi.org/10.1111/j.1745-3992.1999.tb00004.x

Stivers, T. (2005). Modified repeats: One method for asserting primacy rights from second position. *Research on Language and Social Interaction, 38*(2), 131-158. https://doi.org/10.1207/s15327973rlsi3802_1

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language & Social Interaction, 41*(1), 31–57. https://doi.org/10.1080/08351810701691123

Strong, M., & Baron, W. (2004). An analysis of mentoring conversations with beginning teachers: Suggestions and responses. *Teaching and Teacher Education, 20*(1), 47-57, https://doi.org/10.1016/j.tate.2003.09.005

Şahin, S. (2019). *An Analysis of English language testing and evaluation course in English language teacher education programs in Turkey: Developing language assessment literacy of pre-service EFL teachers* [Unpublished doctoral dissertation]. Middle East Technical University.

Taras, M. (2002). Summative and formative assessment: Perceptions and assessment. *Active Learning in Higher Education*, *9*(2), 172-192. https://doi.org/10.1177/469874080991655

Tasker, T. Q., & Herrenkohl, L. R. (2016). Using peer feedback to improve students' scientific inquiry. *Journal of Science Teacher Education*, *27*(1), 35–59. https://doi.org/10.1007/s10972-016-9454-7

Taşpınar, M., & Halat, E. (2009). Yeni ilköğretim 6. sınıf matematik programının ölçme değerlendirme kısmının öğrenci görüşleri doğrultusunda incelenmesi.

Uludağ Üniversitesi Eğitim Fakültesi Dergisi, 22(2), 551-572. Retrieved from https://dergipark.org.tr/tr/pub/uefad/issue/16690/173459

Ten Have, P. (2007). *Doing conversation analysis: A practical guide* (2nd ed.). SAGE Publications Ltd.

Topping, K. (2009). Peer assessment. *Theory Into Practice*, *48*(1), 20-27. https://doi.org/10.1080/00405840802577569

Topping, K. (2017). Peer assessment: Learning by judging and discussing the work of others. *Interdisciplinary Education and Psychology*, *1*(1), 1-17. http://dx.doi.org/10.31532/InterdiscipEducPsychol.1.1.007

Topping, K. (2021). Digital hardware for peer assessment in K-12 schools and universities. *Frontiers in Education*, *6*. https://doi.org/10.3389/feduc.2021.666538

Torrance, H. (1993). *The nature of creativity as manifest testing.* Press Syndicate of the University of Cambridge.

Vásquez, C. (2004). "Very carefully managed": Advice and suggestions in post-observation meetings. *Linguistics and Education, 15*(1-2), 33–58. https://doi.org/10.1016/j.linged.2004.10.004

Vickerman, P. (2007). Student perspectives on formative peer assessment: an attempt to deep end learning? *Assessment & Evaluation in Higher Education*, *2,* 221-230. https://doi.org/10.1080/02602930801955986

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, *11*(4), 374–402. https://doi.org/10.1080/15434303.2014.960046

Vogt, K., & Tsagari, D. (2017). Assessment literacy of foreign language teachers around Europe: Research, challanges and future prospect. *Language Testing and Assessment*, *6*(1), 41-63.

Walsh, S. (2002). Construction or obstruction: teacher talk and learner involvement in the EFL classroom. *Language Teaching Research, 6*(1), 3–23. https://doi.org/1191/1362168802lr095oa

Walsh, S. (2006). *Investigating classroom discourse*. Routledge. https://doi.org/10.4324/9780203015711

Walsh, S. (2011). *Exploring classroom discourse: Language in action*. Routledge.

Walsh, S., & Mann, S. (2015). Doing reflective practice: A data-led way forward. *ELT Journal*, *69*(4), 351–362. https://doi.org/10.1093/elt/ccv018

Waring, H. Z. (2003). Also' as a discourse marker: Its use in disjunctive and disaffiliative environments. *Discourse Studies*, *5*(3), 415–436. https://doi.org/10.1177/14614456030053006

Waring, H. Z. (2007a). The multi-functionality of accounts in advice giving. *Journal of Sociolinguistics, 11*(3), 367–391. https://doi.org/10.1111/j.1467-9841.2007.00328.x

Waring, H. Z. (2007b). Complex advice acceptance as a resource for managing asymmetries. *Text & Talk – An Interdisciplinary Journal of Language, Discourse Communication Studies, 27*(1), 107-137. https://doi.org/10.1515/text.2007.005

Waring, H. Z. (2008). Using explicit positive assessment in the language classroom: IRF, feedback, and learning opportunities. *The Modern Language Journal*, *92,* 577-594. http://dx.doi.org/10.1111/j.1540-4781.2008.00788.x

Waring, H. Z. (2011). Learner initiatives and learning opportunities in the language classroom. *Classroom Discourse*, *2*(2), 201-218, https://doi.org/10.1080/19463014.2011.614053

Weng, F., & Shen, B. (2022). Language assessment literacy of teachers. *Frontiers in Psychology*, *13*, 864582. https://doi.org/10.3389/fpsyg.2022.864582

Xu, Y. (2017). Assessment literacy of university English teachers in China: A mixed methods study. https://doi.org/10.5353/th_991043979535703414

Xu, Y., & Brown, G. T. (2017). University English teacher assessment literacy: A survey-test report from China, 6(1), 133-158. https://researchspace.auckland.ac.nz/docs/uoa-docs/rights.htm

Yan, X., Zhang, C., & Fan, J. (2018). 'Assessment knowledge is important but…: How contextual and experiantial factors mediate assessment practice and training needs of language teachers. *System*, *74,* 158-168. http://dx.doi.org/10.1016/j.system.2018.03.003

Yastıbaş, A. E., & Takkaç, M. (2018). Yabancı dilde ölçme değerlendirme okuryazarlığının gelişimini anlamak. *Bingöl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *8*(15). https://doi.org/10.29029/busbed.364195

Yeşilçınar, S., & Kartal, G. (2020). EFL teachers' assessment literacy of young learners: Findings from a small-scale study. *Kuramsal Eğitimbilim Dergisi [Journal of Theoretical Educational Science], 13*(3), 548-563.

You, H. J. (2014). *Checking recognition: Do you remember and do you know in talk-in-interaction* [Unpublished doctoral dissertation]. University of Illinois, IL.

You, H. J. (2015). Reference to shared past events and memories. *Journal of Pragmatics*, *87,* 238-250. https://doi.org/10.1016/j.pragma.2015.02.003

Yorke, M. (2005). Formative assessment in higher education: Its significance for employability, and steps towards its enhancement. *Tertiary Education and Management, 11*(3), 219-238, https://doi.org/10.1007/s11233-005-5110-z

Yöney, R. Z. (2021). *Assuming Learner Behaviour as an interactional resource in L2 testing and evaluation classroom interaction in a teacher education context*. [Unpublished master's dissertation]. Hacettepe University.

Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching, 49*(4), 461-493. doi:10.1017/S0261444816000161

Yüksel, İ. (2011). The effects of post-observational feedback modes on teaching beliefs: Peer vs. Teacher-mediated feedback. *Turkish Online Journal of Qualitative Inquiry, 2(1),38-56.*

Yüksel, İ., & Başaran, B. (2020). Reflective peer feedback in the practicum: Qualitative and quantitative practices. *Turkish Online Journal of Qualitative Inquiry*, *11*(1), 85-109. https://doi.org/10.17569/tojqi.627310

Zeichner, K. (2006). Studying teacher education programs: Enriching and enlarging the inquiry. In C. Condrad, & E. Serlin (Eds.), *The handbook for research in education* (pp.78–94). SAGE Publications.

Zhu, Q., & Carless, D. (2018). Dialogue within peer feedback processes: clarification and negotiation of meaning. *Higher Education Research & Development*, *37*(4), 883-897. https://doi.org/10.1080/07294360.2018.1446417

APPENDICES

A. JEFFERSON TRANSCRIPTION CONVENTION (2004)

Symbol	Definition and use	Key(s)
[yeah] [okay]	Overlapping talk	
=	End of one TCU and beginning of next begin with no gap/pause in between (sometimes a slight overlap if there is speaker change). Can also be used when TCU continues on new line in transcript	
(.)	Brief interval, usually between 0.08 and 0.2 seconds	
(1.4)	Time (in absolute seconds) between end of a word and beginning of next. Alternative method: "none-one-thousand-two-one-thousand...": 0.2, 0.5, 0.7, 1.0 seconds, etc.	
Word Wo:rd	Underlining indicates emphasis Placement indicates which syllable(s) are emphasised Placement within word may also indicate timing/direction of pitch movement (later underlining may indicate location of pitch movement)	
wo::rd	Colon indicates prolonged vowel or consonant One or two colons common, three or more colons only in extreme cases.	
↑word ↓word	Marked shift in pitch, up (↑) or down (↓). Double arrows can be used with extreme pitch shifts.	↑ Wingdings 3 (104) ↓ Wingdings 3 (105) ↑ ALT+24 ↓ ALT+25
.,_¿?	Markers of final pitch direction at TCU boundary: Final falling intonation (.) Slight rising intonation (,) Level/flat intonation (_) Medium (falling-)rising intonation (¿) (a dip and a rise) Sharp rising intonation (?)	¿ ALT+168
WORD	Upper case indicates syllables or words louder than surrounding speech by the same speaker	
°word°	Degree sign indicate syllables or words distinctly quieter than surrounding speech by the same speaker	° ALT+248
<word	Pre-positioned left carat indicates a hurried start of a word, typically at TCU beginning	
word-	A dash indicates a cut-off. In phonetic terms this is typically a glottal stop	
>word<	Right/left carats indicate increased speaking rate (speeding up)	
<word>	Left/right carats indicate decreased speaking rate (slowing down)	
.hhh	Inbreath. Three letters indicate 'normal' duration. Longer or shorter inbreaths indicated with fewer or more letters.	
hhh whhord	Outbreath. Three letters indicate 'normal' duration. Longer or shorter inbreaths indicated with fewer or more letters. Can also indicate aspiration/breathiness if within a word (not laughter)	
w(h)ord	Indicates abrupt spurts of breathiness, as in laughing while talking	
£word£	Pound sign indicates smiley voice, or suppressed laughter	
#word#	Hash sign indicates creaky voice	
~word~	Tilde sign indicates shaky voice (as in crying)	
(word)	Parentheses indicate uncertain word; no plausible candidate if empty	
(( ))	Double parentheses contain analyst comments or descriptions	

B. MONDADA MULTIMODAL TRANSCRIPTION CONVENTION (2018)

* *	Gestures and descriptions of embodied actions are delimited between
+ +	two identical symbols (one symbol per participant)
Δ Δ	and are synchronized with corresponding stretches of talk.
*--->	The action described continues across subsequent lines
---->*	until the same symbol is reached.
>>	The action described begins before the excerpt's beginning.
--->>	The action described continues after the excerpt's end.
.....	Action's preparation.
----	Action's apex is reached and maintained.
,,,,,	Action's retraction.
ric	Participant doing the embodied action is identified when (s)he is not the speaker.
fig	The exact moment at which a screen shot has been taken
#	is indicated with a specific symbol showing its position within the turn at talk.

C. ETHICS COMMITTEE APPROVAL

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59

Sayı: 28620816 / 332

04 EKİM 2019

Konu: Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Doç.Dr. Çiler HATİPOĞLU, Nilüfer Can DAŞKIN

"İngilizce Öğretmen Adaylarının Ölçme ve Değerlendirme bilgi ve becerilerine yönelik gelişimlerinin sınıf-içi etkileşim içinde izlenmesi" başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülmüş ve **332 ODTÜ 2019** protokol numarası ile onaylanmıştır.

Saygılarımızla bilgilerinize sunarız.

[redacted]
Prof. Dr. Tülin GENÇÖZ
Başkan

[redacted]
Prof. Dr. Tolga CAN
Üye

Doç.Dr. Pınar KAYGAN
Üye

[redacted]
Dr. Öğr. Üyesi Ali Emre TURGUT
Üye

[redacted]
Dr. Öğr. Üyesi Şerife SEVİNÇ
Üye

[redacted]
Dr. Öğr. Üyesi Müge GÜNDÜZ
Üye

Dr. Öğr. Üyesi Süreyya Özcan KABASAKAL
Üye

[redacted]

D. TURKISH SUMMARY / TÜRKÇE ÖZET

Bir Öğretmen Eğitimi Bağlamında Gerçekleşen Yabancı Dilde Ölçme ve Değerlendirme Sınıf İçi Etkileşiminde Etkileşimsel Bir Kaynak Olarak Test İlke ve Prensiplerine Referans Gösterme

Bu çalışma İngilizce öğretmenliği son sınıf öğrencilerinin İngiliz Dili Öğretimi programının müfredatında zorunlu bir ders olan İkinci dilde ölçme ve değerlendirme dersinde hazırladıkları İngilizce testleri üzerine gerçekleştirdikleri akran geribildirimi etkileşimlerine odaklanmaktadır. Mevcut literatür, ölçme ve değerlendirme faaliyetlerini eksiksiz ve kusursuz şekilde yönetebilme kabiliyetinin öğretmenlerin işlerinin en önemli parçalarından biri olduğunu vurgulamakta ve öğretmenlerin mesleki hayatlarının neredeyse yarısını oluşturduğunu göstermektedir. Öğretmenlerin ölçme ve değerlendirme faaliyetlerine zaman ayırmasındaki en önemli husus, içinde bulundukları eğitim programlarının ihtiyaç ve beklentilerine cevap verebilmek ve bu doğrultuda verdikleri eğitimin kalitesini değerlendirerek öğrenci, veli ve kuruma yönelik geribildirim sağlayabilmektir. Sonuç olarak içinde bulunduğumuz 21.yüzyılda öğretmenleri kendilerinden beklenen ve mesleki yeterliliklerinin bir parçası olan ölçme ve değerlendirme faaliyetlerini en iyi şekilde hazırlama ve yürütebilme yetisini karşılayabilir nitelikle yetiştirmek öğretmen eğitim programlarının en önemli hedeflerinden biri olmuştur. Fakat önceki çalışmalar vurgulayarak belirtmiştir ki yabancı dilde ölçme ve değerlendirme öğretmen eğitimi programlarındaki öğretmen adaylarını ve hizmet içindeki öğretmenleri bile zorlamasıyla bilinen edinimi en zor olan alanlardan biridir. Bunun en temel sebeplerinden biri öğretmen eğitimi programlarında verilen yetersiz düzeyde hazırlanmış yabancı dilde ölçme ve değerlendirme dersleri ve hizmet içindeki yetersiz öğretmen eğitimi programlarıdır. Lisans düzeyindeki verilen ölçme ve değerlendirme derslerin büyük bir çoğunluğu geleneksel yöntemle hazırlanmış ve teorik ders konularına (test türleri, dil yeteneklerini farklı türdeki testlerle ölçme vb.) odaklanmıştır. Derslerin müfredatları

genel olarak teorik konulara odaklanırken öğrencileri aslında mesleki hayatlarında birçok kez karşılaşacakları test hazırlama ve test maddesi yazma gibi konularda gerekli pratiği yapmaktan alı koymaktadır. Bunun başlıca sebepleri, öğretmen eğitimi programlarındaki personel yetersizliği, kalabalık öğrenci nüfusu ve ders sayısının yetersizliğinden kaynaklanmaktadır. Bu bağlamda, uluslararası düzeyde ve Türkiye'de yürütülen birçok çalışma, öğretmenlerin yabancı dilde ölçme ve değerlendirme okuryazarlığının çok düşük seviyelerde olduğunu ve bunun dolayısıyla İngilizce eğitimini olumsuz yönde etkilediğini ortaya çıkarmıştır. İngilizce öğretmenlerinin düşük seviyedeki ölçme ve değerlendirme okur yazarlıkları onları çalıştıkları kurumda geleneksel yöntemlere dayalı içinde bulundukları eğitim bağlamının ihtiyaçlarını karşılayamayan ve bu yüzden hali-hazırda bulunan İngilizce sınavlarını kullanmaya itmektedir. Fakat önceki çalışmaların birçoğu belirtmiştir ki, öğretmenlerin geleneksel test yöntemlerini benimseme tutumları onları 21.yüzyıldaki yüksek öğretim hedeflerinin gerektirdiği alternatif değerlendirme yöntemlerini uygulamaktan mahrum bırakmıştır. Alternatif değerlendirme yöntemlerinin başında gelen öz ve akran değerlendirmesi dünyadaki American Eğitimciler Derneği (AAE) ve Türkiyedeki Yükseköğretim Kurulu (YÖK) gibi önemli heyetlerden eğitim kurumlarında benimsenmesi ve sıklıkla uygulanmasına yönelik resmi destek ve teşvik almıştır. Fakat öğretmenlerin bu gibi yeni alternatif yöntemlerini içselleştirerek benimsemeleri için ilk etapta bunu lisans dönemlerinde öğrenmeleri ve mesleki süreçlerinin doğal bir parçası haline getirmeleri gerekmektedir. Birçok çalışmaya göre, bu gibi yeni metotların benimsenmesi için mezuniyetten sonraki hizmet içi eğitim çok geç kalınmış ve verimsiz sonuçlar doğurmaktadır. Bu yüzden alandaki mevcut birçok araştırma öğretmen eğitimi programlarına odaklanmıştır. Bunların başında gelen çalışmalar genel olarak öğretmen bilişi başlığı altında öğretmenlerin programlara yönelik ve ilaveten ölçme ve değerlendirmedeki algı, inanç ve değerlerine ilişkin çalışmalar yürütmüştür. Fakat eğitim programlarındaki öğretmen eğitimi dersleri genel olarak öğretim ilke ve yaklaşımları, dil becerilerinin öğretimi vb. temelde 'nasıl öğretmeli' sorusuna odaklanırken 'nasıl ölçmeli' sorusuna gerekilen önem verilmemiştir. Bunda başlıca sebep birçok devlet ve özel üniversitedeki dil eğitimi öğretmen yetiştirme programlarındaki ders müfredatlarının çoğunluğunda yabancı dilde ölçme ve değerlendirme dersinin gereken ilgiyi görmemesi ve sadece bir dönemde teorik olarak sunulmasıdır. Önceki çalışmalar bu yönde öğretmenlerin ders

hakkında görüş ve değerlerine ses vermiş ve sonuç olarak ölçme ve değerlendirme dersinin ‘nasıl ölçmeli’ sorusuna yanıt verir nitelikte yapılandırılması gerektiği kanısına varmıştır. Bu bağlamda akademisyenler yabancı dilde ölçme ve değerlendirme dersine yönelik öğretmen ve öğretmen adaylarının fikirlerine önem verirken, alanda ‘öğretmen adaylarının ölçme ve değerlendirme okuryazarlığını nasıl edindiklerine ve ölçme-değerlendirme yetisine nasıl kazandıklarına dair herhangi bir çalışma yürütülmemiştir. Oysa yabancı dilde ölçme ve değerlendirme öğretmen adaylarının pedagojik disiplin bilgilerinin önemli bir parçası olmasıyla birlikte gelecekti mesleki hayatlarının da temelini oluşturan bir bilgi alanıdır. Bu sebeplerden ötürü öğretmenlerin yabancı dilde ölçme ve değerlendirmeye yönelik bilişsel tutumlarının yanı sıra sınıf içindeki etkileşimleri göz önünde bulundurularak bu alandaki gerekli yetkinliği nasıl geliştirdiklerine yönelik bir çalışma bu tez sayesinde alana sunulacaktır. Geçmiş çalışmalar göstermiştir ki öğretmenlerin mesleki hayatlarında uyguladıkları öğretim yöntemleri, sınıf içindeki uygulamalarının çoğunluğu öğretmenlerin öğrencilik hayatlarındaki tecrübe ve sosyal etkileşimlerinden beslenerek şekillenen ‘öğretmen bilişlerinin’ bir parçasıdır. Buna göre, bu çalışmanın ana odak noktası olan 4.sınıf İngilizce öğretmen adaylarının yabancı dilde ölçme ve değerlendirme dersindeki sınıf içi sosyal etkileşimlerine odaklanarak ölçme ve değerlendirme okuryazarlığını nasıl edindiklerine yönelik bir çalışma yürüterek, alandaki ‘nasıl ölçmeli’ sorusuna bir cevap oluşturacaktır. Öğretmen eğitimi alanında yürütülen önceki çalışmalar göstermiştir ki mevcut çalışmaların birçoğu nitel ve nicel olarak yürütülmüş ve öğretmenlerin sosyal etkileşimlerine yeterince odaklanılmamıştır. Oysaki öğretmen eğitimi çerçevesini yeniden kavramsallaştıran ünlü çalışmalarında Freeman ve Johnson (1998) öğretmenleri “öğretme öğrenenler” ve “ingilizce dil öğretimini anlama ve geliştirmenin merkezi" olarak etiketlemektedir.” Buna paralel olarak, bu çalışma, öğretmen adaylarının sınıf içi etkileşimsel yeterliliklerinin (Walsh) gelişimini “anlama aracı haline gelen” konuşma-içi etkileşimlerini yerinde yakından gözlemleyerek ve analiz ederek öğretmen adaylarının yaşadıkları öğrenme süreci hakkında daha derin bir anlayış kazanmayı amaçlamaktadır (Hall, 2003, 2011). Bu sebepten bu çalışmada, Ankaradaki bir İngiliz dili eğitimi programına kayıtlı 4.sınıf öğretmen adaylarının ter-yüz edilmiş sınıf modelinde ölçme ve değerlendirme dersi için hazırlamış oldukları İngilizce sınavlarına akran geribildirimi (peer feedback) verirken girdikleri

etkileşimler odak noktası olmuştur. Bu bağlamda çalışma mevcut literatürden ‘söylem çözümlemesi yöntemi ile’ ayrılmaktadır. Ayrıca yazarın bilgisi kapsamında, daha önce ölçme ve değerlendirme dersi Türkiyedeki öğretmen eğitimi programlarının hiçbirinde “ters-yüz edilmiş sınıf”(flipped-classroom) modelinde verilmemiştir. Sonuç itibarıyla, çalışma hem kendine özgü bağlamı hem de öğrencileri akran geri bildirimi alternatif yöntemiyle ölçme ve değerlendirme dersinde teori ve pratiği entegre edilmesiyle ortaya çıkan sosyal etkileşime odaklanmaktadır. Bu sosyal etkileşimleri en ince ayrıntısıyla çalışarak İngiliz dili öğretmen adaylarının yabancı dilde ölçme ve değerlendirme okuryazarlığını nasıl edindiklerine ve geliştirdiklerine dair bir anlayış geliştirmeye katkıda bulunmayı hedeflemektedir. Çalışmanın odak noktasını oluşturan İngiliz dili öğretmen adayları Ankara'da bir devlet üniversitesi yabancı dil eğitimi Bölümü'nde son sınıfa kayıtlı 23 öğretmen adayıdır. Bu çalışmadaki öğretmen adaylarından on yedisi kadın, 6'sı ise erkektir. Anadilleri Türkçedir ve yetkin İngilizce konuşanlardır. Avrupa Birliği dil portfolyosuna (CEFR) göre lisans eğitimine başlayabilmek ve üniversitelerinde İngilizce hazırlık programından muaf olabilmek için öğrencilerin en az 70 (TOEFL 86 ve IELTS 7.0'a eşit) başarı notu ile İngilizce yeterlik sınavını geçmeleri gerekmektedir. Öğretmen adayları çalışmaları süresince eğitim programları olan yani ELT müfredatlarına göre test ve değerlendirme ile ilgili herhangi bir ders almamışlardır. Bu nedenle katılımcılar, İngilizce dil eğitimi programının son sınıf müfredatında zorunlu temel ders olarak yabancı “dilde ölçme ve değerlendirme” (ELTE) dersini almışlardır. Ders, son sınıf öğretmen adaylarına yılda iki defa olmak üzere farklı zaman aralıklarında verilmektedir. 2018-2019 eğitim-öğretim yılı güz döneminde 14 hafta süren kurs, yaz okulunda ise altı hafta süren yoğun bir kurs olarak okutuldu.

Hatipoğlu’ nun (2015,2017) öne sürdüğü gibi, ölçme ve değerlendirme dersi ELT müfredatının ayrılmaz bir parçasıdır. İngilizce eğitim programlarında verilen dersler, öğretmen adaylarına gerekli becerileri kazandırmayı amaçlamaktadır. Öğretmen adaylarının dil değerlendirme yeterliliğini ve okuryazarlığını geliştirmeye yönelik bu standartlar ışığında, bu çalışma ters çevrilmiş sınıf modelini (flipped-classroom) benimseyen 2018-2019 yaz okulu dönemine odaklanmıştır.

Ters çevrilmiş sınıf modeli temel olarak öğrencilerin ders okumalarını ve ödevlerini evde tamamlayıp öğrenme etkinliklerine katılmak üzere sınıfa hazırlıklı gelecekleri şekilde tasarlanmıştır. Bishop ve Verleger'in (2013) özetlediği gibi "dersler ve ödevler sınıf dışında tamamlanabilirken, aktif öğrenme etkinlikleri sınıf içinde gerçekleşmektedir." Bu çalışma bağlamında akran geribildiriminin şu şekilde uygulanmıştır. Dersi veren öğretmen adayları ve öğretim görevlisi ile ELTE haftada iki kez pazartesi günleri 3 saat, Salı günleri ise 4 saat görüşmüştür. Dersin ilk iki buçuk haftasında test ve değerlendirmenin tanıtılması, farklı uluslararası ve ulusal test türleri, test yapımı ve madde yazma süreci, güvenilirlik ve geçerlilik ölçümleri konularında geniş bir konu yelpazesi ele alınmıştır. Bu arada öğretmen adayları sınavlarının ilk taslaklarını yazmak üzere akran grupları oluşturdular. Bishop ve Verleger'e (2012) göre, akran geribildirimini uygulamadan önce atılması gereken en kritik adımlardan biri öğrencileri hazırlık sürecine dahil etmektir. Bu nedenle, grupların tasarımı, sınavların sunulması ve geri bildirim oturumları için genel şema, ELTE kursuna katılan öğretmen adaylarıyla tam işbirliği içinde kararlaştırılmıştır.

Sınavlar 5,6,7 ve 8. sınıf öğrencileri için hazırlanmış olup 6 ana bölümden (kelime bilgisi, dilbilgisi, okuma, yazma, dinleme ve konuşma) oluşmuştur. Ders, 3. hafta itibariyle, öğretmen adaylarının ödevlerini evde tamamladıkları, sınavlarını gruplar halinde hazırladıkları ve üniversitenin ODTUCLASS adı verilen çevrimiçi öğrenme yönetim sistemindeki akranlarının sınavları için yazılı geri bildirim sağladıkları ters çevrilmiş sınıf modelini hayata geçirdi. Her grup sırasıyla iki akran grubu için geri bildirim sağladı ve gruplar alınan geri bildirimlere göre sınavlarını düzeltti. Ancak, güz döneminde 14 hafta boyunca verilen kapsamlı dersin aksine, öğretmen adaylarının aldıkları geri bildirimlerden en iyi şekilde yararlanmaları için sınırlı bir süreleri vardı. Tüm grupların akranları tarafından yazılan sınavlara ilişkin yorumlarını içerik, biçim, pratiklik ve geçerlilik açısından sunmaları beklenmektedir. Buna ek olarak, öğretmen adayları, ders profesörünün yabancı dil testi, değerlendirmesi ve değerlendirmesine dayanan kitap ve makalelerden oluşan atanmış ders okumaları yoluyla yorumlarını doğrulamak zorunda kaldılar. Öğretmen adaylarının akranlarından aldıkları geri bildirimlerin ardından sınavlarında gerekli revizyonları yapmaları ve revize edilmiş versiyonlarını 6. Haftadan önce ODTUCLASS'A yüklemeleri beklenmiştir. 6. Haftaya

gelindiğinde öğrenciler bir sunum oturumu gerçekleştirdiler ve sınavlarının son halini tüm sınıfa tanıttılar ve öz değerlendirme yaparak sınav yapım sürecine yansıttılar.

Çalışmaya yalnızca verilerin geri bildirim ve sunum oturumlarından alınan kayıtlar dahil edildi. Veri setinin tamamı kayıt altına alınmış olsa da verilerin ana odak noktası yaz döneminde verilen dersin toplam 12 saat 728 dakikasını oluşturan 5 dersten ibarettir. Bu bağlamda, doğal konuşma verisinden oluşan bu çalışmayı analiz etmek için 'söylem çözümlemesi' (conversation analysis) yöntemi uygulanmıştır. Bu yöntemin ana metotlarından ilki doğal olarak ortaya çıkan konuşma verisini analiz edilebilir hale getirebilmek için video kayıtlarının çevriyazın (transcription) formatına uyarlanmasıdır. Çevriyazın sistemi Seedhouse'un (2005) belirttiği gibi," transkriptler kaçınılmaz olarak eksiksiz, okunabilirlik ve kapsamlılık arasında her zaman bir uzlaşma içeren birincil verilerin seçici olarak işlenmesidir" (s. 251). Bu nedenle, 2001 yılında Chris Fassnacht tarafından nitel verileri yönetmek için geliştirilen bilgisayar destekli transkripsiyon yazılımı programı Transana aracılığıyla, araştırmacılara yorucu manuel transkripsiyon yöntemi yerine daha yönetilebilir bir platform sağlamak için 12 saatlik veriler ortografik olarak çevriyazın sistemine uyarlanmıştır. Daha sonra öğrencilerin geri bildirim etkileşimi oturumlarında uyguladıkları olası bir olgunun keşfi için, Gail Jefferson (1984, 2004) ve Mondada'nın (2018) jestler, yüz ve vücut hareketleri gibi somutlaşmış eylemlerin entegrasyonu ile bilinen çevriyazın sistemi tarafından geliştirilen yaygın olarak kullanılan transkripsiyon simgeleri uygulanmıştır. Transkripsiyonlar analiz için verilerin ayrıntılı bir temsilini sunduktan sonra (Sert, 2015), verileri tekrar tekrar izlenmiş ve transkriptleri analiz için söylem çözümlemesinin içeriden bakış açısı yöntemiyle (emic) okunmuştur. İçeriden bakış açısı perspektifi önceden belirlenmiş herhangi bir varsayımı reddeder ve yalnızca verinin kendisine bağlıdır ve (Wong & Waring, 2010) arkasındaki mantığı şöyle tanımlar, "İçeriden bakış açısı perspektifi, dile ve sosyal etkileşime içeriden bir bakış açısıyla bakmanın, yani kendini katılımcıların yerine koyarak konuşmalarını ve eylemlerini anlamaktır." Verileri birden çok kez yakından inceleyerek ve bir odak noktası olmadan (unmotivated looking) bakış sınırları içindeki ayrıntılı transkripsiyonlardan geçerek, etkileşimciler tarafından normatif olarak yönlendirilen olası bir fenomen kanıta dayalı olarak keşfedilir. Bunun sonucunda, öğretmen adaylarının, akranlarına geribildirim sağlarken yorumlarını "test ilkelerine atıfta

bulunarak” (RTP) destekledikleri ve kanıt getirdikleri keşfedilmiştir. Öğretmen adaylarının yaşıtları tarafından tasarlanan testlerdeki sorunlara dikkat çekerken ‘test ilkelerine atıfta bulunma’ olgusu analitik olarak teşhis edilmiştir. Olgu, test ilkelerine ilişkin kurallara atıfta bulunurken gereklilik kipi '-meli/-malı ' yapısı kullanılarak keşfedilmiş olup 12 saatlik verinin içerisinden toplanan 77 örnek (extract) ile bir örneklem koleksiyonu oluşturulmuştur. Bulunan bu fenomene örnek olarak bir öğretmen adayı bir akranın sınavındaki test sorularının yapılandırılmasına ilişkin geri-dönüt sağlarken “soruya ve şıklara bakarak seçenekleri elimine etmemeliyiz” ilkesine atıfta bulunarak negatif bir yorumda bulunmuştur. İlaveten görülmüştür ki, yaz okulunun son haftalarına doğru, test ilkelerine öğretmen adayları tarafından daha terimsel bir dil kullanımıyla atıfta bulunulmuştur. Örneğin, “ilk soru elimine edilebiliyor şıklardan, bu nedenle soruda problem var” gibi. Bu, ilkelerin zaman geçtikçe yaz okulunun sonlarına doğru öğretmen adaylarının yabancı dilde ölçme ve değerlendirme okuryazarlıklarına eklenmiş olabileceği bu yüzden terimsel olarak kolaylıkla ifade edilebildiğine yorulabilir.

Keşfedilen fenomenin (RTP) öğretmen adaylarının etkileşimlerinin anlık analizi doğrultusunda, aşağıdaki araştırma sorularını yanıtlaması amaçlamaktadır:

1) Akran geri bildirim etkileşiminde RTP'nin ortaya çıktığı dizisel bağlamlar nelerdir?

2) RTP'nin akran geribildirim etkileşiminde gerçekleştirdiği temel işlevler nelerdir?

3) Öğretmen adaylarının geribildirim sağlarken atıfta bulunduğu test ilkeleri nelerdir?

4) RTP öğretmen adaylarına ölçme ve değerlendirme becerileri için nasıl öğrenme olanakları ve fırsatları sağlamaktadır?

Bu araştırma sorularının öncülüğünde yapılan analiz göstermiştir ki ‘test ilkelerine atıfta bulunma” fenomeni üç farklı dizisel bağlamda ortaya çıkmaktadır. Bunlardan ilki öğrenciler test maddelerindeki problemlere kanıt getirirken, ikincisi bulunan problemlere test ilkelerine uyan ve ihlal etmeyen önerilerde bulunulurken, üçüncüsü yapılan negatif yorumlara karşı gelinirken. Sonuçlar göstermiştir ki akran geribildirimi etkileşiminin ilk aşaması, öğretmen adaylarının test maddelerini sorunsallaştırması,

yani sınavların test ilkelerini ihlal etmesi ve dil sınavlarının geçerliliği ve güvenilirliğinde potansiyel sorunlar yaratmasıdır. İkinci aşama, etkileşimli bir kaynak olarak test ilkelerine yönelerek, dil sınavlarının sorunlarına karşı savunulması sırasında geri bildirim veren gruplar ile test-yazar grupları arasında direnişin ortaya çıkmasına odaklanmaktadır. Etkileşimin son aşaması, öğretmen adaylarının test ilkelerine atıfta bulunmalarına odaklanırken, belirlenen sorunları ortadan kaldırmak ve test ilkelerinin yerine getirilmesi gerektiği perspektifinden iddialarını sunarak dil testlerinin kalitesini artırmak için sorunlu test öğelerine olası bir çözüm önermektedir. Kurumsal bağlama benzer şekilde, bu araştırmada, yabancı dilde ölçme ve değerlendirmenin teorik ve pratik yanlarını bütünleştirmek için ters çevrilmiş bir sınıf modelini benimseyen öğretmen adaylarının sınıf etkileşiminin de, 'pedagojik görev' yani akran geribildirimi çerçevesinde şekillenen, yinelenen değişim aşamalarına da sahip olduğu görülmektedir.

Fenomenlerin oryaya çıktığı dizisel bağlamlara odaklanırsak, ilk ortaya çıktığı ‘test maddelerindeki sorunlara atıfta bulunma’ aşamasında çoğunlukla akran geri bildirim grupları tarafından başlatılan sorunsallaştırma aşamasında akran geri bildirim etkileşiminde ortaya çıkmaktadır. Analiz, 77 RTP örneği arasında, sorunlara atıfta bulunma kategorisinden 38 örnek elde edildiğini göstermiştir. Öğretmen adaylarının odak noktasındaki test maddelerini değerlendirmeleri sırasında RTP'nin etkileşimsel bir kaynak olarak sıkça kullanılması, iddialarını veya iddialarını destekleme zorunluluğundan kaynaklanabilir. Antaki ve Leudar'a (1990) göre, mevcut taraflar arasında doğal olarak tartışmalı olan iddiaları sunarken, öğretmen adayları ortaya çıkabilecek direniş ile başa çıkabilmek için iddialarını sağlam gerekçelerle desteklemeye ihtiyaç duymaktadırlar. Bununla birlikte, test ilkelerine yönelmek sadece iddiaları destekleme güdümlü değildir. Bunun yerine, öğretmen adaylarının, ELTE dersinin ilk haftalarındaki teorik derslerden sunulan ölçme ve değerlendirmedeki ortak bilgiyi ortaya çıkarmak için üstlendikleri bir çalışmadır. Bu nedenle, bu ilkelere atıfta bulunurken, öğretmen adayları, kullanımlarında açıkça görüldüğü gibi, bu test ilkelerinin kaynağına tanınabilir referanslar sağlar (hocam siz bize söylediniz, Heaton diyor ki , Robert diyor). Bu, etkileşime getirilen bu test ilkelerinin epistemik kaynağının esas olarak öğretmen adaylarının dil değerlendirme okuryazarlıklarını oluşturdukları paylaşılan dış kitap kaynaklarından ve geçmiş

öğrenmelerden ve dersi öğreten hocadan (Can Daşkın, 2017; Can Daşkın & Hatipoğlu, 2019; Siz, 2015) kaynaklandığı anlamına gelir. Öğretmen adaylarının test ilkelerini kullanırken bu paylaşılan kaynaklara başvurarak bir test öğesindeki sorunu dile getirmeleri "öğretmen adaylarının geçerli ve güvenilir bir yabancı dil testin nasıl olması gerektiğine ilişkin anlayışlarını" gösterebilir. Bu sayede açıklanan bir soruna kanıt getirmek için test ilkelerini kullanmaları, akran geri bildirim gruplarını ortaya çıkan sorunları fark etmeye ve sorunun doğası hakkında karşılıklı ortak bir anlayış ve kabullenişe yönlendirir.

Fenomenin ortaya çıktığı ikinci aşama olan 'geri bildirime karşı direnç gösterme' test öğesindeki sorunun duyurulması, sorunun kabulü veya test madde-yazar grupları tarafından reddedilmesi olarak ilgili ikinci eylemi yapar. Bu nedenle, akran geri bildirim etkileşimindeki ikinci aşama, akran geri bildirim grupları tarafından sağlanan ani bir olumsuz geri bildirim karşısında sınavlarını savunmak için madde yazar gruplarının başlattığı direnç aşamasıdır ve bu da daha uzun bir müzakere ve dolayısıyla odak test öğesi için karar verme süreci ile sonuçlanır. Başlangıç olarak, öğe yazar gruplarının sağlanan olumsuz geri bildirime yönelimli olarak karşı koydukları 11 direnç örneği çalışmanın ana koleksiyonundan elde edilmiştir. Ancak, her direniş test-maddesinin sorunsallaştırılmasının karşısında geçerli bir argüman olarak kabul edilmemiştir. Bu nedenle, bazıları sınavlarını savunmak için geçerli itirazlar sunarken, bazıları geçerli bulunmayıp reddedilmiş ve test ilkelerinin kapsamında ele alınmıştır. Bu nedenle, akranlar arası geri bildirim etkileşiminin direnç aşamasında RTP'nin kullanılması, test-maddesi grupları ve / veya akranlar arası geri bildirim grupları tarafından başlatılan direnç aşamasına bir çözüm getirmek için bir "problem çözme mekanizması" işlevi görmüştür. Bu anlamda çalışmamız Waring (2017) tarafından yapılan çalışmadan farklıdır. Çalışmasında Waring, "eğitimsel ilkelerine atıfta bulunarak mentorların tavsiye verme stratejilerini inceledi. Daha geniş çaptaki eğitimsel ilkelere başvurmanın taraflar arasında karşılıklı anlayışın sağlanmasına yardımcı olduğunu keşfetti. Ayrıca, eğitimsel ilkeler, mentor ile öğretmen arasında gerginlik ve anlaşmazlık yaratmada öğretmenlerin veya öğretmen adaylarının uyum sağlamasına yardımcı olmuştur. Çalışmamızda, ortaya çıkan anlaşmazlıklarda test ilkelerinin çağrılması "problem çözme aracı" olarak işlev görmektedir. Okul öncesi

öğretmenleri ilkeleri yönlendirerek iddialarını haklı çıkarır ve devam eden anlaşmazlıklara çözüm getirir.

Akran geribildirim etkileşiminin son aşamasında, öğretmen adayları, test öğesinin geçerliliği ve güvenilirliğine yönelik bir tehdit olan sorunu ortadan kaldırmak için işbirliği içerisinde bir değişiklik önermektedir.

Bu nedenle öğretmen adaylarının farklı dizisel ortamlarda RTP kullanarak önerilerini formüle ettikleri alıntılarda görülmektedir. Test ilkelerine yapılan atıflar, yorumların derhal kabul edilmesinin önünü açtığından öğretmen adayları, önerilerini test ilkelerinden kanıtları önce (pre-account) veya sonraki (post-account) pozisyonlarda "kanıt/ açıklama "sunarak test madde-yazar gruplarına destekleyerek sunmuşlardır.

Önerilere RTP'nin eşlik ettiği ikinci ardışık ortam, sunulan önerilerin uygunluğunun ve geçerliliğinin değerlendirilme aşamasıdır. Test maddesinde ortaya çıkan sorunun ortadan kaldırılmasına yönelik önerinin açıklanmasından sonra öğretmen adaylarının önerilerini değerlendirdiğini ve önerinin çürütülebilirliğine kanıt getirmek için test ilkelerine yöneldiğini, bunun da başka bir test ilkesini ihlal ettiğini ve dolayısıyla test öğesinde ek sorunlara yol açtığını göstermektedir. Bu nedenle, teklif edilen öneriye itiraz etmek, maddede ortaya çıkabilecek ek sorunların önceden belirlenmesine katkıda bulunur ve bu nedenle önerinin iptal edilmesine yol açar, bu da test ilkelerine uygun makul başka bir çözüm önerisinin aranmasına yol açar.

Bütün bu dizisel bağlamlara bakıldığında "test ilkelerine atıfta bulunma" fenomeninin ilk etapta test maddelerine olumsuz bir geribildirim verirken, test maddesindeki soruna kanıt getirme işlevini gerçekleştirdiği görülmüştür. İkincil olarak geribildirime direniş gösterirken test maddesi-yazar ve geribildirim sağlayan gruplar tarafından problem-çözme ve direnişle başa çıkma işlevini yerine getirdiği gözlemlenmiştir. Üçüncü aşama olan test maddelerindeki sorunları ortadan kaldırmak için önerilerde bulunurken ise, yapılan önerilerin ölçme ve değerlendirme ilkelerine uygunluğu ve geçerliliği kapsamında değerlendirildiği bu sayede açıklama yapma/destekleyici kanıt getirme işlevlerinde bulunduğu tespit edilmiştir.

Çalışmanın analizi göstermiştir ki öğretmen adaylarının ağırlıklı olarak farklı dizisel ortamların ihtiyaçlarına göre birçok farklı çeşitte test ilkelerine yönelmişlerdir. Bunlar genel olarak test maddesi yazma ve yapılandırma ilkeleri, test içeriklerine ve test maddelerinin zorluk derecelerine yönelik test ilkelerinden oluşmaktadırlar. Dördüncü araştırma sorusuna ithafen ise, test-ilkelerine atıfta bulunma olgusunun ters-yüz edilmiş sınıf ortamında 'akran geribildirimi' yöntemi sayesinde öğrencileri daha anlamlı diyaloglara teşvik ettiği, bu sayede yabancı dilde ölçme ve değerlendirme dersinde zengin bir öğrenme ortamı sağladığı tespit edilmiştir. Akran geri bildirim etkileşimi, öğrencilere birbirlerinin performansı hakkında güncel ve zamanında ayrıca spesifik, detaylı bir geri bildirim sağlamak için yapıcı tartışmalara katkıda bulunduğu gözlemlenmiştir. Bu çok katılımlı (multilogue) etkileşimler, sorunlu test öğeleri üzerine detaylı düşünmeye, sorunları tespit etmeye ve dolayısıyla sorunlu dil testlerinde iyileştirmelerle sonuçlanmıştır. Bu anlamda akran geribildirimi etkileşimi, öğretmen adaylarının geribildirimleri almalarında ve epistemik durumlarındaki öğrenme değişimlerinde açıkça görüldüğü üzere "akran öğreniminin" potansiyelini ortaya koymuş ve bu da dil testlerinin maddesel yapılarında iyileşmelere yol açmıştır. Öğretmen eğitimi bağlamında akran geribildiriminin uygulanmasının öğretmen adaylarını geçerli test öğeleri oluşturma performanslarına olumlu yansıyan zayıf ve güçlü yönlerini müzakere etmeye yönlendirdiğini söyleyebiliriz. Amacı doğrultusunda, akran geribildirimi, öğretmen yansıtma uygulamalarında sıklıkla ihmal edilen yapıcı ve zamanında geri bildirim sağlayarak öğrencilerin performansına katkıda bulunmuştur. Bu çalışmanın literatüre olan ikinci olası sonucu ise RTP olgusu ile ortaya çıkmaktadır. Öğretmen adaylarının geribildirim oturumların ilk haftalarındaki test ilkelerine yönelimleri, ELTE dersi boyunca öğretmen adayları için bir kılavuz oluşturan kuralların birlikte oluşturulmasına neden olmuştur. Bu anlamda RTP, öğretmen adaylarına, aksi takdirde fark edilmeyecek olan test maddelerindeki olası sorunların tespiti konusunda özel rehberlik ve yönergeler sağlamıştır. Ayrıca, RTPler etkileşimin ihtiyaçlarına göre şekillendirilmiş ve ortaya çıkmıştır. Bu da dil testlerinin iyileştirilmesine ve ölçme ve değerlendirmenin bir parçası olan madde yazma ve gözden geçirme becerilerinin geliştirilmesine neden olmuştur.

Özetlemek gerekirse, bu çalışmanın en önemli katkısı, öğretmen adaylarının geri bildirim etkileşimlerinin, gelecekteki işlerinin önemli bir parçası olan ancak yabancı

dilde ölçme ve değerlendirme dersinin yapılandırılmasında çoğunlukla ihmal edilen test-maddesi yazma ve gözden geçirme uygulamalarında onlara ‘uygulamalı bir deneyim’ (Köksal, 2014) sağlamasıdır. Önceki araştırmalar, “teorinin genellikle pratiği bilgilendiremediği" sorununu ortaya çıkarırken (Johnson, 1996, sg.766) öğretmen eğitimi programlarına bu derslerin yapılandırılmasına ilişkin (Şahin & Subaşı, 2019), bu çalışma, öğretmen adaylarının ölçme-değerlendirme becerilerini geliştirmek ve onlara yardımcı olmak için “diyalojik, yansıtıcı uygulamalara" katılmalarını sağlayan bir model ortaya koymaktadır.

Ayrıca, bu çalışma öğretmen eğitimi programlarında ELTE kursuna ayrılan ders saatlerinin artırılmasını önermektedir. Ders genellikle sadece bir dönem için sunulduğundan (Şahin, 2019) çoğu konu kapsamlı bir şekilde ele alınmamaktadır. Bu sebepten bu çalışma dersin alternatif bir yöntem olan ters-yüz edilmiş sınıf içi modelinde gerekli koşullar sağladığında öğretmen adaylarına pratiği ve teoriyi nasıl entegre edebileceklerine ilişkin başarılı bir model sunmaktadır.

E. THESIS PERMISSION FORM / TEZ İZİN FORMU

ENSTİTÜ / INSTITUTE

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences ☐

Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences ☒

Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics ☐

Enformatik Enstitüsü / Graduate School of Informatics ☐

Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences ☐

YAZARIN / AUTHOR

Soyadı / Surname : BAYDAR
Adı / Name : MÜBERREM BERNA
Bölümü / Department : English Language Teaching

TEZİN ADI / TITLE OF THE THESIS (**İngilizce** / English):

Reference to Testing Principles as an Interactional Resource in L2 Testing and Evaluation Classroom Interaction in Teacher Education Context

TEZİN TÜRÜ / DEGREE**:** **Yüksek Lisans** / Master ☒ **Doktora** / PhD ☐

1. **Tezin tamamı dünya çapında erişime açılacaktır. /** Release the entire work immediately for access worldwide. ☒
2. **Tez iki yıl süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of **two years. *** ☐
3. **Tez altı ay süreyle erişime kapalı olacaktır.** / Secure the entire work for period of **six months**. ***** ☐

***** *Enstitü Yönetim Kurulu kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir. / A copy of the decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.*

Yazarın imzası / Signature .....**Tarih** / Date .....
(Kütüphaneye teslim ettiğiniz tarih. Elle doldurulacaktır.)
(Library submission date. Please fill out by hand.)