Australian Journal of Teacher Education

Manuscript 6182

Developing Pre-service Teachers' Assessment Literacy in the Practicum: An Action Research Study

Mustafa Akın Güngör
Gazi University, College of Foreign Languages, Türkiye
Müzeyyen Nazlı Güngör
Gazi University, Gazi Faculty of Education, Türkiye

Abstract: This paper reports on an action research (AR) study that pre-service teachers conducted to assess their students in practicum. Based on Xu and Brown's (2016) assessment literacy (TALIP) framework, we aim to empower them in a real classroom atmosphere. Data were collected systematically and came from interviews, pre-service teachers' own exams, observations, evaluations of students, and reflections. The qualitative analysis was used to identify the realities and constraints of the classroom, to plan alternative ways of assessment, to reveal the observation results, and to make sense of the reflections. Results showed that pre-service teachers gained an awareness of the gap between the EFL curriculum and the assessment practices in real classrooms, built self-confidence over their engagement with AR for preparing to assess, and developed a tacit understanding of assessment literacy through experience in real context. Suggestions are provided for assessment courses in pre-service teacher education, practicum practices, and future research.

Keywords: assessment literacy; action research; pre-service; practicum; teacher education

Introduction

Research on pre-service teachers (PSTs) learning about assessment for their practice as teachers has increasingly gained popularity in teacher education programmes (DeLuca et al., 2013; Looney et al., 2018). Despite the growing interest, studies have shown that PSTs are unprepared and feel incompetent in assessment practice (Poth, 2013; Xu & He, 2019). Among the reasons for this are factors such as limited time (usually one semester) allocated for the assessment component providing a general introduction to assessment (Xu & Brown, 2016), lack of opportunities to provide PSTs with assessment practices in classroom contexts (Siegel & Wissehr, 2011), and a lack of research on assessment pedagogies (Mertler, 2004) in teacher education are identified.

When these PSTs start the profession in countries where there is a strong testing tradition such as China (Chen & Brown, 2018), Hong Kong (Lam, 2015), and Turkiye (Güngör & Güngör, 2021), they are mostly placed within tensions and conflicts between using high-stakes and summative exams for selection and ranking purposes, and alternative and formative assessment types featured in the K-12 English as a foreign language (EFL) curriculum. PSTs are, on the one hand, expected to develop assessment knowledge, identify test types, prepare assessment tasks and tests, and enact their assessment practices in real or

simulated contexts (Cheng et al., 2010). On the other hand, when they become teachers, they are asked to prepare students for assessment of learning through summative and high-stakes exams. While PSTs are educated theoretically for test types, test preparation, and alternative assessment methods at the pre-service teacher education (PSTE) level, they face the complex practical nature of summative and high-stakes exams in real contexts. Thus, the assessment component does not effectively prepare PSTs for assessment practices in real contexts due to a lack of opportunities to enable them to observe in-class assessment realities to develop their communication, reflection, and decision-making skills, and implementation of their own assessment ways in PSTE (Looney et al., 2018; Poth, 2013).

Addressing this criticism, this study aims to empower PSTs to implement assessment in a real classroom atmosphere through action research (AR). PSTs conducted an AR study in their practicum context to identify students' and mentors' assessment-related problems, to design and implement appropriate assessment tasks for learning, and to reflect on the implementation of their tasks. The paper reports on their AR processes and practices during the development of their assessment literacy. It also aims to understand assessment-related problems in real classrooms, engage PSTs in decision-making and problem-solving stages in practicum, evaluate students' exams, and generate feedback for their future learning.

Pre-service Teachers' Assessment Literacy Development

Assessment literacy (AL) is a concept first presented by Stiggins (1991) and broadly defined as teachers' assessment capabilities to plan and implement appropriate assessment tasks, and to use assessment results for communication and improvement of instruction and student learning (Looney et al. 2018). Recently, the concept has been extended beyond the AL to new models (Abell & Siegel, 2011; DeLuca, 2012) and frameworks (Xu & Brown, 2016) that comprehensively include all teacher education and development phases.

Xu and Brown's (2016) TALIP framework provides a new, contextualised, and practice-based perspective on AL. There are six components of their conceptualisation: a) knowledge base, b) teacher conceptions of assessment, c) institutional and socio-cultural contexts, d) TALIP as the main concept, e) teacher learning, and f) teacher identity reconstruction as assessors.

The knowledge base covers the knowledge of disciplinary and pedagogical content, assessment aims, content and methods, grading, feedback, peer- and self-assessment, assessment interpretation and communication, and assessment ethics. These necessary aspects are usually the focus of most assessment courses similar to those of pre-service English language teacher education programmes in Turkiye (CoHE, 2018), yet are decontextualized and standardised assessment knowledge. It is argued that the uptake of this knowledge occurs through teacher conceptions which include teacher beliefs about the nature and aims of assessment including PSTs' cognitive and affective dimensions (Lutovac & Flores, 2022). Hence, this second component refers to what teachers believe to be true or false about assessment, the degree of consistency between the new knowledge and the existing practices, and their positive or negative experiences and emotions which may lead them to resist or accept changes for new assessment uses and policies. In countries such as China and Turkiye, PSTs are selected and appointed to their new universities and preferred programmes as a result of high-stakes national exams. As students, they came from a strong testing tradition in which they were taught to the test and were tested summatively at high stakes throughout their schooling. Unlike the Finnish PSTs whose assessment conceptions varied from assessment-positive to assessment-cautious and assessment-critical in Kyttälä et al.'s (2022) study, these PSTs start the assessment course in the programme with previous

conceptions, most of which come from the negative, troublesome, and mechanical experiences of learning and being tested. Throughout the process of assessment education, their conceptions importantly change over practicum (Xu & He, 2019) or may remain unchanged (Deneen & Brown, 2016).

The teacher conception is followed by institutional and socio-cultural realities which necessitate context-bound goals and outcomes. Xu and Brown (2016) categorise these contextual realities under micro- and macro-contexts. The former refers to teachers' immediate workplace actions, social norms, and dominant discourses in classrooms while the latter refers to national assessment policies. As an example, English language teachers in Turkiye are to follow the K-12 EFL curriculum (MoNE, 2018) in their teaching and assessment. Renewed in 2018, the curriculum is dominated by fundamental perspectives and principles in language teaching: English as an international language, communicative competence, integration of four skills, collaboration and learner autonomy, and alternative and authentic assessment methods. Specifically, these teachers are encouraged to use alternative assessment methods such as portfolios, task-based projects, and observation-based techniques (communicative activities) for K-12 grades to maintain consistency between the teaching and assessment dimensions; however, they are not trained as to how these methods and tools can be designed and implemented in diverse contexts. Besides, the national evaluation system in Turkiye rests on high-stakes exams with multiple-choice types focusing only on reading comprehension, vocabulary, and grammar knowledge. These teachers find themselves in disequilibrium (Güngör & Güngör, 2021) as a result of the strict boundaries set by the national system and classroom teaching practices, and the tensions arising from this incongruence.

Xu and Brown (2016) define the TALIP, the fourth component, as "the art of compromise" in which teachers have to "balance the demands of external factors and constraints with their own beliefs and values" (p. 157). It is a combination of teachers' assessment knowledge, conceptions, and responses to contextual realities and needs. Being aware of classroom dynamics such as preparedness of students, age-appropriate assessment tasks, course materials, and classroom atmosphere, and making better compromises in their decision-making and enacting through reflection and involvement in communities of practice. Therefore, teacher learning - the fifth component - is seen as the impetus for TALIP advancement. Finally, when these teachers understand their roles as assessors (Looney et al., 2018) and transform their identities from being students who were assessed and tested into teachers who plan and implement assessment tools, evaluate students, and communicate feedback to students for improvement, their assessment identity - the last component - begins to emerge, flourish, and shape through interactions with stakeholders.

Studies (Abell & Siegel, 2011; DeLuca et al., 2013; Hill et al., 2014; Siegel & Wissehr, 2011; Xu & Brown, 2016) call for increased opportunities for improving PSTs' AL through a connection between the theory and practice of assessment. The interplay between PSTs' past beliefs, assessment histories, conceptions, and their emerging beliefs on assessment throughout the programme, their negotiation with stakeholders, and their contextual assessment experiences have become the focus of attention for future studies.

The Role of Practicum in Developing PSTs' Assessment Literacy

Drawing from the PSTs' need for acquiring assessment knowledge in the field through reflection and collaboration about classroom experiences as mentioned in these studies, we designed an action research study as part of the practicum for PSTs who were taking the assessment course for the first time in the teacher education programme. Practicum

engages PSTs with supervised teaching experience, systematic observation, and gaining familiarity with a specific teaching setting. With the shift from a positivist to an interpretivist perspective in second language teacher education (Johnson, 2009), learning to teach has been recognised as a social, contextualised, dynamic, and cognitive process. Accordingly, PSTs come to understand their knowledge, adapt and use that knowledge in contexts where they teach, and make sense of their practices over time.

Aligned with the sociocultural views of teacher education (Johnson, 2009), the nature of AL has also been conceptualised as dynamic, context-laden, situational, culturally responsive, and social (Willis et al., 2013). DeLuca and Klinger (2010), and Xu and He (2019) recognise the practicum as an important and convenient context for PSTs to develop their AL, conceptions of and confidence in assessment, arguing that assessment courses only are not sufficient in developing PSTs' assessment confidence and skills and that without understanding the classroom complexities teachers' AL may not truly develop. PSTs need to make observations, critically examine mentors' assessment practices, identify problems, negotiate with stakeholders, develop their own solutions, and test them in real classrooms. DeLuca and Klinger (2010) cautioned that PSTs' reliance on mentors may not lead to a general and consistent AL development because the knowledge and skills of mentors may be incomplete, outdated, and biassed. Therefore, the embodiment of critical and reflective practices and perspectives is required for the quality of PSTs' AL development in practicum (Lam, 2015).

Recent studies have drawn attention to the content quality of practicum for AL development opportunities. To be specific, Giraldo and Murcia (2019), in their action research study with PSTs in a Colombian teacher education programme, concluded that practicum provided contextual problem-based scenarios and an interactive atmosphere for AL development of PSTs. In their 'Improving Formative Assessment in Vocational Education and Adult Literacy, Language and Numeracy Programmes' (IFA) project where they used problem-based methodology as an action research strategy, Swann et al. (2011) also highlighted the power of this methodology to enable teachers to change their practice for the better by enhancing their understanding of what they can and cannot do in their workplace. Wyatt-Smith et al. (2017) highlighted research-informed learning opportunities for PSTs in practicum, discussing the necessity of collecting student learning data for use in monitoring student progress, adapting assessment methods for students' future learning improvement, and determining how to use the data for their own AL development. Lam (2015) examined the content of assessment-related courses in five different teacher education programmes in Hong Kong. As a result, a list of suggestions was made for high-quality assessment education: a) sociocultural and sociopolitical dimensions of assessment should be emphasised in assessment education since they are the determining factors that affect teaching and learning in the classroom, b) assessment courses should be updated with recent research, materials, and on-the-job opportunities for workshops, c) PSTs' knowledge base for using classroom-based assessment results to inform large-scale assessment results for students' improvement should be developed and space must be given for testing this knowledge in practicum, d) teacher educators must assess PSTs through loop input in which they also include AL into the assessment of PSTs' pedagogical skills for enhanced awareness and competence, and e) since AL is impossible to be observed in one lesson, PSTs may be encouraged to video record and/or audio record their assessment practices in practicum for formative and summative purposes.

Although the aforementioned studies and others (Brevik et al., 2017; McGee & Colby, 2014) enabled PSTs to assess real students' work on written texts or video-/audio-recorded performances, there is a dearth of research to understand how PSTs critically reflect on their observations and understandings regarding classroom assessment practices, how they

relate classroom realities (micro-context) with national assessment system (macro-context), and how they make decisions to assess their students concerning these realities. For this reason, the current study aims to answer the following research questions:

1) What are the assessment-related problems PSTs identified in real classrooms?
2) How do PSTs make decisions to solve these problems?
3) What are the reflections of PSTs on the action research process?

Methodology

Research Context and Participants

This study reports on a cyclical action research model based on Kemmis and McTaggart (as cited in Burns, 2010). PSTs needed to understand the language teaching and assessment experiences of mentors and students so that they could determine a parallel assessment practice in practicum for their final assessment in the "testing and assessment in ELT" course of the ELT programme. As suggested by the lecturer in the post-conference seminars, they followed the AR model and went through four phases of AR: plan, act, observe, and reflect iteratively.

We employed a case study (Dörnyei, 2007) which enabled us to focus on the unique complexity of the practicum, draw conclusions from PSTs' experiences and reflections in this unique context, and transfer them to other teacher education programmes. The second author was teaching "testing and assessment in ELT" which was offered in the fourth year of the programme while the first author managed data collection and analysis. The content of this course involved a general introduction to testing and assessment, assessment types, factors affecting assessment, traditional and contemporary assessment, sample tests and assessment tasks, and evaluation.

Concurrent with this course, PSTs were also doing their practicum at a state school in Central Anatolia in the 2021-2022 academic year. Two PSTs (pseudonyms: Yağmur and Can) agreed to participate in this study voluntarily, gave consents, and then were matched with two different mentors at two different state schools. Primary school refers to education from 1st to 4th grades (7-10 years), secondary school from 5th to 8th grades (11-14 years), and high school from 9th to 12th grades (15-18 years) in Turkiye. Yağmur attended a 7th grade class at a secondary state school with one mentor while Can attended an 11th grade class at a high school with another mentor. All classes had approximately 30 students. Their practicum school placement was organised by the supervisor. They were at the age of 22 and were in their fourth year of the English Language Teaching Programme at a large state university in Turkiye.

The Council of Higher Education (CoHE, 2018) in Turkiye categorises teacher education under three basic knowledge and competence fields: subject matter (48%), pedagogical and pedagogical content knowledge (34%). In addition to these fields, world knowledge courses are also offered and constitute 18% of the teacher education curriculum. Accordingly, PSTs in Turkiye take subject matter courses such as advanced reading, writing, speaking and listening skills, translation studies, linguistics, pedagogical courses such as educational sciences, inclusive education, counselling, classroom management, and pedagogical content knowledge courses such as approaches and methods in ELT, teaching English to young learners, materials adaptation, testing and assessment in ELT, and practicum (https://www.yok.gov.tr/Documents/Kurumsal/egitim_ogretim_dairesi/Yeni-Ogretmen-Yetistirme-Lisans-Programlari/Ingilizce_Ogretmenligi_Lisans_Programi.pdf).

They had already been taught assessment principles, types, and methods (the first component "knowledge base" in TALIP) when they started AR study in practicum. Hence,

practicum would provide them with an authentic experience to transfer what they learnt in theory into practice by challenging them to use their decision-making, problem-solving, and critical thinking strategies. While they were tested through a pen and paper exam for the midterm in the ELT programme, as part of the final assessment PSTs were asked to design and implement an exam in the real context, evaluate the exam, communicate the results to students in practicum, and plan their next step. The AR study took PSTs two months (May and June 2022) to complete the cyclical process.

Data Analysis

Inductive data analysis was used for all the data sets. As Mackey and Gass (2005) noted, "... the goal is generally for research findings to emerge from the frequent, dominant, or significant themes within the raw data, without imposing restraints as is the case with predetermined coding or analysis schemes" (p. 179). Without any predetermined codes, we began analysing the data first on our own. We read and scanned the data several times until we had identified significant themes under the main categories by following research questions. Then, we had 25% of them coded by each other for interrater reliability. In total, we analysed four self-originated exams, four reflection papers, two mentors' and a group of students' interview results before and after the assessment practice, and PSTs' data analysis. We achieved 95% agreement on the 25% of the data from each set.

Action Research Procedure

As outlined in their final assignment, the PSTs were to apply a self-designed exam in their practicum class. It was PSTs' responsibility to determine the specifications, focus, skills, and content of the exam. Notably, we were in our first face-to-face education year after two years of global lockdown. In the post-conferences, mentors and students reported on their teaching, learning, and assessment process during the pandemic, comparing them with the face-to-face conditions, and trying to get used to the new normal at the time of the study. When PSTs were assigned to the assessment task, they questioned what kind of an exam it should be, what contents and objectives they should include, and the implementation method. In other words, PSTs needed to understand the mentor's assessment preferences and students' proficiency levels, their previous assessment histories and ideas, and identify classroom realities and constraints in their practicum contexts. They thought their answers would help them prepare for age-appropriate exams which would also be parallel to their practice teaching methods and content. In an attempt to identify these points, PSTs in the first phase developed two interview forms (Tab.1), one for the mentor and one for the student.

Mentors	students
1. What language skills were assessed in the pandemic? 2. How were they assessed? 3. What do you think about your assessment practices? 4. Is there any change about assessment in the post-pandemic?	1. How were you assessed in the pandemic? 2. What do you think about your assessment experiences in the pandemic? 3. What kind of differences are there between pandemic and post-pandemic assessment?

Table 1. Interview questions for mentors and students

Each PST in collaboration with the assessment lecturer analysed the answers to determine the situation. They felt that their answers alone would not help them to portray the situation, so they decided to analyse the secondary school curriculum, course books, and previous exams. Overall, PSTs' data in this phase came from interviews with mentors and students, secondary and high school EFL curricula, course books, and previous exams. This took one week to complete. The data were analysed in the post-observation sessions in which they exchanged their data and shared ideas on each other's practicum context with the lecturer in the next week. PSTs' were also asked for their reflections on the data collection and their own interpretation of the assessment in schools. In other words, two weeks were spared for data collection and analysis to identify the assessment situation and design an appropriate task.

During the second phase, PSTs compared and contrasted curriculum guidelines, learning objectives, and classroom teaching practices to determine the content and type of assessment they would implement. They referred to practicum observation notes which they kept in previous practicum tasks to examine past lessons. Drawing on these data, their assessment conceptions, and the theoretical background they gained in the course, they first prepared their test specifications. Next, these specifications were shared and discussed with peers and the supervisor in the post conferences where they gave feedback on each other's specifications and assessment plans. After the revisions, they finalised the exam and prepared their rubrics. Then, they applied the exams in real classes and evaluated them. They noted down the challenges they experienced in the application.

The third phase consisted of mentors' feedback, student interviews for reflection, and communicating the exam results to students so that they could observe the effects of their self-prepared exams and document the opinions of those involved. For this reason, they used interview questions for mentors and students, and exam results as data. These data were to be used for PSTs' next round of assessment practices such as test specification and exam preparation, rubric evaluation, and feedback giving.

In the fourth phase, PSTs reflected on their AR process from the first stage to this. Their reflections occurred as a group over the post-conference sessions with the supervisor and were categorised under four parts as also suggested in Burns (2010): reflections on assessment practice, on the AR process, and feelings and experiences. Next, they documented all the assessment and AR files and data and submitted them to the lecturer for their own evaluation in the course.

Results

Case 1 – Yağmur

Yağmur was a silent, hardworking, and motivated pre-service teacher, attended secondary school practicum. In her interview with the mentor, she stated that although grading students was easier during the pandemic, maintaining their attention during the lessons was very tiring. She argued that online education was only beneficial when students' readiness level was higher and motivation was up. She uncovered:

> *The mentor was not happy with remote education. She sent photos of exam sheets to students via WhatsApp and collected them back in the same way. The exam was on vocabulary and grammar. Students answered them in their rooms. They were evaluated on exam results and regular attendance in the online lessons. This was a problematic situation because most students' cameras were not working properly. So, the teacher could not see their faces. Also, their houses were noisy, most of the time, so they reported that they could not concentrate. Some students felt sleepy and bored due to the lack of eye contact and interaction. They became more and more unsociable, which affected their psychology negatively.*

Analysing the exams mentors used in face-to-face education, Yağmur realised that the exam format and content were more or less the same as that of remote education though the exam delivery system changed during the pandemic. In other words, technology (WhatsApp or Web 2.0 tools) was used to test students' structure and vocabulary knowledge through mechanical and traditional exams. No listening or speaking part was included, or no formative assessment was preferred. Students' regular attendance was evaluated along with the exam results. She understood that old habits die hard. Despite the curriculum guidelines which suggest the use of formative and alternative assessment methods through communicative tasks, mentors continued traditional teaching and assessment methods. One student in her class stated:

> *Because there was no exam during the pandemic, there was no regular attendance. We did homework on our notebooks and sent their photos to the teacher. It was an exhausting term for me because I was stressed too much when I could not open my camera or attend classes synchronously. But now we are being tested again. I could not benefit from the pandemic term, so I have difficulty catching up with listening, speaking, and writing activities in the lesson.*

Having examined the course book and learning objectives of the 7th graders' second term content (A2 level - 6th unit) as part of the one practicum task before, Yağmur knew that students were expected to listen, read, write, interact, and speak about suggestions, arrangements and action sequences (MoNE, 2018). She had prepared her teaching practice lessons in a way to improve these skills and functions of the language. Having combined the information she gained from mentors, students, and document analysis and followed the exam preparation steps in the course, she first prepared a test specification. She aimed to outline target language skills and components along with learning objectives and exam details (duration, delivery, number of items, test type, and scoring). Next, she had one of her peers and then her mentor check her test specifications. Her peer focused on the variety of test types to monitor students' language use in four skills. However, her mentor criticised the fact that the number of reading comprehension questions was not enough for checking reading comprehension and she had missed one main idea question for listening comprehension. Yağmur revised her specifications and added new items. In one of the post-conferences, she raised some of her concerns:

Although I learnt assessment theory and how to prepare test specifications and exams, I am having trouble with finding authentic reading and listening materials to be used in the exam. I am also hesitant about the scoring part for each language skill. What if I demotivate students?

With those hesitations in mind, she prepared her first draft of an integrated language exam consisting of language functions and structure smoothly embedded in the language skills items. Her mentor checked the first draft in terms of the difficulty of items for students while her peer focused on the smooth integration of language skills, topic selection for language skills, and the match between learning objectives and test types. Yağmur revised the exam and shared the second draft with her lecturer. She corrected several typing errors, clarified some of the instructions, and added duration in some parts. Finally, her exam was ready. Wary of students' language proficiency and teaching practices, she carefully prepared rubrics and answer keys. In rubric preparation for speaking and writing skills, she could solely make use of CEFR guidelines (CEFR, 2021) and curriculum objectives because her mentor had not used any rubric before. This would help her design realistic and age-appropriate expectations of the students. Before the exam, her mentor warned her about possible discipline problems. On the day of the implementation, she was nervous about students' possible unpredictable reactions toward her and the exam. She managed the class and implemented the exam in two hours.

After the exam, some students said, "We have never taken such an exam before". Speaking skills were the most problematic element for them because they did not understand her instructions or questions in English. Yağmur knew the fact that her mentor usually translated every utterance into Turkish (L1-native language) to ease comprehension and not to lose time in lessons; nevertheless, she believed that she should not use L1 translation to challenge students to use English. Unfortunately, her choice of L2 use turned the speaking exam into chaos. As students did not try to understand her, they made noise and asked repetitive comprehension questions.

I was sure of my speaking skills because, in my previous teaching practices, I did various speaking practices with the class. As I tried to explain speaking items and role-plays, students asked more and more questions without listening to me. They were nervous too since it was their first time to take such an exam.

In scoring, she saw some students give alternative answers that were not mentioned in the key, so she accepted them. She found scoring productive skills easier than receptive skills due to the rubrics.

Rubrics made it easy to understand students' speaking and writing performances. I believe I scored them fairly thanks to the rubrics.

In communicating the exam results, she shared scores with each student individually and had some counselling sessions for developing their language skills more effectively, referring to learning strategies and sharing some of her favourite techniques.

It was in the form of an informal conversation. I gave feedback, let them review their exam sheets, and shared my suggestions without demotivating them. Now, I understand what assessment is!

Case 2 - Can

Can was a motivated, hardworking, and extroverted pre-service teacher, attending a high school practicum. High schools had slightly different regulations in the pandemic in terms of assessment implementation. Contrary to secondary school students, high school students went to school only to take the exam during the pandemic. At the beginning of the

pandemic - in emergency remote teaching - they were all at home, attending classes online. The only assessment way his mentor used was:

> *I assessed students in three ways: online attendance, comprehension questions during the online lessons, and face-to-face exams at school. In my online classroom, students were expected to attend at least 11 weeks of the lessons out of 14 weeks. During their attendance, I asked each student comprehension questions individually to see whether they followed me and understood the topics. After the emergency remote teaching, students came to schools with masks on their faces to take the exams.*

Although the mentor said that she followed the curriculum properly and taught the topics, she followed the same teaching techniques used in the pre-pandemic. She did not change the techniques or revise them according to online education realities or students' needs. Also, since there was no exam at first, students did not find a reason for studying. For these reasons, students became unmotivated, attended less and less, and lost their attention towards the end of the term. From the students' perspective,

> *Our feelings about the assessment are negative. We sometimes could not attend the classes due to internet or technical problems. When we did, our teacher was teaching English as they did in face-to-face education. There were no speaking or listening practices. They just taught grammar, presented vocabulary asked comprehension questions, and used exams like they did face-to-face. Since we could not benefit from these lessons, the assessment type was not suitable for our needs and situations.*

The problem with students' assessment practices was due to their teacher's teaching techniques and classroom management. Can realised that the mentor could not update her teaching techniques following online education conditions and technological developments. Using the same techniques with the same course book (only via sharing the pages of the book on the screen) did not hold their attention. Most of the students could not find a way to listen, read, speak, and write in the lessons. As a result, the productive skills and receptive skills of students did not improve at all. Can also added:

> *In the face-to-face context, students practised all skills but reading was more emphasised. The mentor did not prepare any lesson plans but followed the curriculum and course book. So, I chose learning objectives, topics, and teaching outcomes from these sources for preparing my assessment.*

Having realised the local constraints and classroom realities, Can compared and contrasted the curriculum guideline, unit objectives, and the mentor's lessons. In his observation of the lessons, he checked whether students could achieve the objectives by interviewing them and observing their performances in the activities. In one of the post-conferences, he reported:

> *I realised the mentor considered classroom participation in the activities in each lesson as a formative assessment tool. Other than this, students were assessed via summative assessment tools such as classical pen and paper exams. There is no spoken interaction or production exam, and no listening assessment is done. The exams start with reading text, comprehension questions, and vocabulary items, and at the end, they are asked to write an opinion paragraph. I know assessing their spoken production and interaction will be difficult*.

Though Can felt such exams were mechanical and trivial, he still valued them as a necessary component of a skills-based communicative exam. Motivated by this belief, Can prepared his lesson plans and applied skills-based activities in his teaching practice before his assessment implementation. These teaching practices aimed to make students familiar with the exam content because Can believed that a language teacher should test what he teaches.

For him, assessment should be parallel to lessons, content, and classroom activities. So, he decided to choose among the topics he taught in the practicum. The first and most challenging part was finding authentic materials to create listening and reading texts appropriate for students' levels.

> *I decided to choose an interview recording from the British Council website and shorten the duration. When writing listening questions, I had the idea of making comics for "fill in the blank" items because I observed my students were interested in manga and comics. Considering their interests, I chose an article on celebrities as a reading text. I prepared one open-ended and seven multiple-choice questions. I adapted the writing task into an e-mail writing one according to the curriculum objective "Students will be able to ask questions to make an interview with a sportsperson.". My writing rubric was also prepared according to this. The speaking task preparation procedure was also the same. I considered CEFR speaking skill categorization, which categorises speaking as production and interaction.*

His first draft of the exam was weak in terms of the wording and content of questions. His peers told him that students may not understand the questions since they contained some colloquial English, some of which may be unknown to them, and some question types were new to students. At the same time, he was also in contact with his mentor throughout the exam preparation process. His mentor supported and counselled him by saying: "I had similar concerns and hesitations at the beginning of my career." She shared one of her previous receptive skills exams so that he could prepare the most appropriate questions for the students. After his feedback from the peer and the mentor, he revised his wording and the content of the questions. In addition, he decided to make a friendly and attractive exam layout because in this process he had the emerging belief that "the more students like an exam, the more they will be enthusiastic and successful."

Getting feedback from peers and mentors and having the lecturer confirm the exam, Can felt secure and finalised the exam. In the implementation process, the students found the exam non-threatening because they became familiar with the exam question types in previous lessons with Can. Their mentor also explained the exam process to the class by emphasising the importance of such skills-based exams for their future career. Although some students were nervous at the beginning of the spoken production exam, they were relieved and enjoyed the topics in the receptive skills (listening and reading).

In the last stage, he assumed evaluating the receptive and productive performances and communicating the results to students would be the easiest part. Contrary to his assumption, especially in the evaluation and communication of the spoken performance, he spent too much time listening to the recording of the students several times, taking notes, and explaining their performance levels to them. This was because he wanted to ensure his intra-rater reliability in evaluating students' spoken performance both during the exam and on the recording after the exam. While giving feedback on their writing, Can used his analytic rubric and correction symbols, which interested students and helped them understand their deficiencies better. In listening and reading evaluation, he used the answer key and let students analyse their papers and his scoring.

Discussion and Conclusion

Based on our results, our discussion mainly focuses on PSTs' assessment-related problems in practicum schools, their decision-making process to address these problems, and their reflections on the AR process. The present AR study provides a qualitative insight into

PSTs' assessment practices and contextual realities in practicum schools by employing a different methodological approach compared to other similar studies. While the AR results shed light on the PSTs' data on their AR journey, we also discovered mentors' and students' experiences and thoughts on assessment from a sociocultural perspective (Johnson, 2009). From this perspective, practicum became a critical component for providing realities to PSTs in this study (Xu & He, 2019).

R.Q.1. What are the assessment-related problems PSTs identified in real classrooms?

The fact that PSTs identified assessment-related problems in practicum schools tells us that students and mentors were also facing these problems. Similar to the stages followed in Giraldo and Murcia's (2019) study, the first stage of the AR "plan" helped PSTs detect these problems, make observations on them, and think about alternative ways to overcome them. We observed that the pandemic affected learning, teaching, and assessment negatively as a whole. Similar to mentors in Lam's (2015) study, mentors in the study felt alone about professional support and transferred traditional face-to-face practices into remote education. This caused many students to participate less in the lessons, lose the sense of meaning in studying for language skills, and become more unsociable. It was also seen that before and after the pandemic, neither formative nor skills-based assessment was implemented although they were suggested in the K-12 EFL curriculum in Turkiye. Mentors were stuck to discrete point tests, and mechanical and traditional assessment ways in language classes (DeLuca, 2012). Due to this fact, students did not have the opportunity to practise and improve their language skills much, which subsequently reduced students' motivation and enthusiasm. Arguably, assessment lecturers' monitoring, guidance, and involvement in this study allowed PSTs to avoid being involved in these idiosyncratic practices.

From PSTs' perspectives, they experienced some problems such as deciding on the exam content, determining the item types according to their assessment beliefs and students' proficiency realities, and finding authentic written and audio texts for reading and listening skills of the exam. They were quite hesitant for fear of choosing overly difficult texts, preparing inappropriate items, and finally, demotivating students to learn English. Although they ensured they improved their assessment literacy, they found it risky and challenging to apply the knowledge into practice in the real context. Since they wrote the items for each skill assessment on their own, they made some mistakes in the wording and content of the questions. In their first attempt to begin such a journey, multiple feedback they received from peers, mentors, and the lecturer facilitated the learning-to-assess process. While lecturer feedback confirmed their assessment beliefs and practices, mentor feedback relieved them by preparing them for classroom realities. Also, peer feedback enabled them to correct their typing and layout errors on the exam paper.

Nevertheless, mentors' L1 translation in classes immediately after the L2 instruction caused habit formation among students, which turned the speaking exam context into chaos in secondary school. Restricted between her beliefs on the insistence of L2 instruction and students' L1 translation desire during the exam - as the PSTs in Lam's study (2015) - Yağmur gave the L1 translation for the sake of managing the class. Students' anxiety affected her negatively too. Since both PSTs felt the necessity to share the evaluation results and communicate them to students in a friendly and supportive role, they also spent too much time in scoring to be holistic and fair. As also noted by DeLuca and Klinger (2010), such a necessity resulting from understanding the complexities of classroom assessment was also derived from their assessment knowledge of "principles of assessment: inter-rater and intra-

rater reliability issues". This social dimension of language assessment emerged as a belief of PSTs (Lam, 2015).

R.Q.2. How do PSTs make decisions to solve these problems?

The "planning" stage in the AR enabled them to experience and observe the problems encountered in the real context. Against these problems, PSTs received support from peers, mentors, and the lecturer in interviews, guidance, and post-observation sessions. From this aspect, they both helped to collect data and to portray the situation in schools (Giraldo & Murcia, 2019). PSTs started to collaborate with their mentors, trust each other, and receive scaffolding from their lecturers. In the exchange of experiences in post-observation sessions, PSTs reminded each other to check the curriculum and course books and shared their assessment knowledge and experiences on previous mentor-made exams. Therefore, it is noted that these sessions helped them gain insight into the assessment problems and produce effective decisions.

In the "action" stage in AR, their collaboration with the mentor and the multiple feedback provided a smooth transition from situation analysis to assessment preparation. The content of this collaboration included PSTs' analysis of curriculum, course books, and lesson plans, their comparisons with the previous teacher-made exams, and their negotiation with mentors on these issues. Due to the positive and constructive nature of mentors, PSTs had a comparatively smooth transition from assessment ideals to assessment reality. Hence, as also argued by Swann et al. (2011), an important finding is that mentor-PST collaboration has a decisive role in PST's establishing assessment beliefs. Interestingly, although mentors guided and helped them to prepare communicative, skills-based exams, they were implementing summative and traditional exams at the time of the study. Their resistance to change or varied assessment methods is worth investigating in future studies.

Although PSTs could observe the effects of the exam on students through feedback and interviews, they did not have the time to make decisions for their future teaching due to the syllabus of the PSTE assessment course. In other words, PSTs were expected to continue with a new AR cycle by revising their exams, using washback effects for their future teaching practices, and designing new assessment tools. Their academic term ended with the "reflection" stage. Thus, they could only reflect on their AR study and assessment practices, which could be attributed as a limitation of this study. Nevertheless, practicum experience with mentor collaboration provided them with the opportunity to conceptualise assessment theoretically and practically (DeLuca et al. 2013). Overall, their decisions were shaped through interviews, multiple feedback, and classroom realities as proposed in TALIP by Xu and Brown (2016).

R.Q.3. What are the reflections of PSTs on the action research process?

PSTs' reflection was categorised into three emerging aspects: reflection on the assessment practice, AR study, and feelings and experiences. To start with the assessment practices, they admitted it was the first time they had ever implemented a skills-based exam for formative purposes. They emphasised the clash they had between contextual realities and assessment knowledge. This was well exemplified in Can's reflection "If we had not had this chance, the tests we have prepared would only be imaginary". In addition, given that our participants were only PSTs, this study is unique in bringing assessment theory and practice to PSTE through AR design. Although she felt stressed and challenged, Yağmur stated this

was the most *authentic* experience in the programme so far, as also named by Wyatt-Smith et al. (2017, p. 262). The excessive planning and precise organisation processes made the assessment practice demanding, she says. They also realised that assessment was not only about preparing and implementing the exam but much more than that. Contrary to Can's expectations, the feedback and communication phases were also demanding for him. As a result, they were more competent and safe in assessment implementation from the beginning to the end.

For the second aspect, it was also their first time to be engaged in an AR design. They found AR stages by Kemmis and McTaggart (as cited in Burns, 2010) smooth and practical since each stage was clearly defined. They could apply their model to the practicum. AR helped them to conceptualise assessment methods in the real context. Can, for example, noted, "Now, I can realise the performance differences among students and how I can assist them with their communication skills." Not surprisingly, both PSTs were satisfied with their assessment process (DeLuca & Klinger, 2010) thanks to the AR study. Therefore, it is notable to carry out studies investigating the relation between PSTs' assessment conceptions and practices in practicum.

As for the last aspect, PSTs named the whole process a rewarding experience that contributed to their assessment identity. They felt devoted to the occupation since their assessment methods were appreciated and used by their mentors in other classes. In the last session, Yağmur was still concerned about the validity of her future exams when she started her profession. So, another important finding of the study is that these PSTs should be traced in terms of their assessment practices in their novicehood, which leads us to a new study. Concurring with Lutovac and Flores (2022), their conceptions, assessment knowledge, and practices could be examined as part of their assessment identity development.

In conclusion, this study contributes to the assessment pedagogy in PSTE by bringing the AR and assessment practice together and presenting a new model, the phases of which were designed, applied, and reflected in a real context. From this aspect, this study addressed the research gap as pointed out by Looney et al. (2018) and Poth (2013). Our focus was on PSTs' assessment practices and AR study only. Hence, the mentors' feelings and practices in the study were not interpreted. In the future, this study should be conducted simultaneously with assessment and practicum components including mentors in PSTE programmes, and outlay the assessment methods for assessing PSTs' assessment practices, as also argued by Wyatt-Smith et al. (2017). One advantage of this would be that PSTs' assessment experiences and methods in diverse practicum contexts can be shared with prospective teachers in PSTE programmes in the form of narratives.

References

Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The professional knowledge base of science teaching* (pp. 205-221). Springer, Dordrecht. https://doi.org/10.1007/978-90-481-3927-9_12

Brevik, L. M., Blikstad-Balas, M., & Engelien, K. L. (2017). Integrating assessment for learning in the teacher education programme at the University of Oslo. *Assessment in Education: Principles, Policy & Practice*, *24*(2), 164-184. https://doi.org/10.1080/0969594X.2016.1239611

Burns, A. (2010). *Doing action research in English language teaching: a guide for practitioners.* Routledge. https://doi.org/10.4324/9780203863466

Chen, J., & Brown, G. T. (2018). Chinese secondary school students' conceptions of assessment and achievement emotions: endorsed purposes lead to positive and negative feelings. *Asia Pacific Journal of Education*, *38*(1), 91-109. https://doi.org/10.1080/02188791.2018.1423951

Cheng, M. M., Tang, S. Y., & Cheng, A. Y. (2012). Practicalising theoretical knowledge in student teachers' professional learning in initial teacher education. *Teaching and Teacher Education*, *28*(6), 781-790. https://doi.org/10.1016/j.tate.2012.02.008

Council of Higher Education (2018). *English Language Teaching Curriculum*, Ankara. https://www.yok.gov.tr/Documents/Kurumsal/egitim_ogretim_dairesi/Yeni-Ogretmen-Yetistirme-Lisans-Programlari/Ingilizce_Ogretmenligi_Lisans_Programi.pdf

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 419-438. https://doi.org/10.1080/0969594X.2010.516643

DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Action in Teacher Education*, *34*(5-6), 576-591. https://doi.org/10.1080/01626620.2012.730347

Deneen, C. C., & Brown, G. T. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education*, *3*(1), 1225380. https://doi.org/10.1080/2331186X.2016.1225380

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

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

Güngör, M. N., & Güngör, M. A. (2021). Reconceptualizing Testing in Times of COVID-19 from an Activity Theory Lens: A Narrative Inquiry of a Test-designer's Experiences. *The Asia-Pacific Education Researcher*, *30*(6), 531-539. https://doi.org/10.1007/s40299-021-00627-0

Johnson, K. E. (2009). *Second language teacher education: A sociocultural perspective*. Routledge. https://doi.org/10.4324/9780203878033

Kyttälä, M., Björn, P. M., Rantamäki, M., Lehesvuori, S., Närhi, V., Aro, M., & Lerkkanen, M. K. (2022). Assessment conceptions of Finnish pre-service teachers. *European Journal of Teacher Education*, 1-19. https://doi.org/10.1080/02619768.2022.2058927

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, *32*(2), 169-197. https://doi.org/10.1177/0265532214554321

Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, *25*(5), 442-467. https://doi.org/10.1080/0969594X.2016.1268090

Lutovac, S., & Flores, M. A. (2022). Conceptions of assessment in pre-service teachers' narratives of students' failure. *Cambridge Journal of Education*, *52*(1), 55-71. https://doi.org/10.1080/0305764X.2021.1935736

Mackey, A., & Gass, S. M. (2015). *Second language research: Methodology and design*. Routledge.

McGee, J., & Colby, S. (2014). Impact of an assessment course on teacher candidates' assessment literacy. *Action in Teacher Education*, *36*(5-6), 522-532. https://doi.org/10.1080/01626620.2014.977753

Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 49-64.

Poth, C. A. (2013). What assessment knowledge and skills do initial teacher education programs address? A Western Canadian perspective. *Alberta Journal of Educational Research*, *58*(4), 634-656.

Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education*, *22*(4), 371-391. https://doi.org/10.1007/s10972-011-9231-6

Stiggins, R. J. (1991). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice*, *10*(1), 7-12. https://doi.org/10.1111/j.1745-3992.1991.tb00171.x

Swann, J., Andrews, I., & Ecclestone, K. (2011). Rolling out and scaling up: the effects of a problem-based approach to developing teachers' assessment practice. *Educational Action Research*, *19*(4), 531-547. https://doi.org/10.1080/09650792.2011.625707

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, *40*(2), 241-256. https://doi.org/10.1007/s13384-013-0089-9

Wyatt-Smith, C., Alexander, C., Fishburn, D., & McMahon, P. (2017). Standards of practice to standards of evidence: Developing assessment capable teachers. *Assessment in Education: Principles, Policy & Practice*, *24*(2), 250-270. https://doi.org/10.1080/0969594X.2016.1228603

Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149-162. https://doi.org/10.1016/j.tate.2016.05.010

Xu, Y., & He, L. (2019). How Pre-service Teachers' Conceptions of Assessment Change Over Practicum: Implications for Teacher Assessment Literacy. *Frontiers in Education*, *4*, 145. https://doi.org/10.3389/feduc.2019.00145

Acknowledgements

We congratulate all our pre-service teachers in acknowledgement of their dedication and hard work during their education.