



# Reconceptualizing Testing in Times of COVID-19 from an Activity Theory Lens: A Narrative Inquiry of a Test-designer's Experiences

Müzeyyen Nazlı Güngör<sup>1</sup> · Mustafa Akın Güngör<sup>2</sup>

Accepted: 29 September 2021 / Published online: 12 October 2021  
© De La Salle University 2021

**Abstract** This article explores the implementation of an online high-stakes language proficiency test during the COVID-19 pandemic through a narrative inquiry of a test-designer, Eda (pseudonym). Situated in the context of Turkish higher education, this study examines Eda's narrative accounts with regard to the development of test adaptation and implementation, technology use, and the emerging contradictions for future language testing practices. The data are analysed based on Clandinin and Connelly's (in Narrative inquiry: Experience and story in qualitative research, Jossey-Bass, 2000) three-dimensional space narrative structure: interaction, continuity, and situation. First, personal and social interaction are analysed to better understand test-designers' dispositions and views, and internal and existential conditions in testing practices in the pandemic. Second, continuity is dealt with in regard to their testing experiences, emerging needs, and implied roles and directions of stakeholders in online testing. Third, the situation is analysed to understand the new landscape where the online test implementation is bounded by pandemic rules and regulations. Third generation activity theory was used to interpret the results of analysis. The study concludes by recognizing how resolutions produce motives for high-stakes test implementation and how the

dynamic relation between constructs of activity systems supports the continuity of online test beyond pandemic times.

**Keywords** Remote testing · L2 proficiency · High-stakes · COVID-19 · Activity theory · Narrative Inquiry · Test-designer

## Introduction

The COVID-19 pandemic caused rapid move from face-to-face to remote teaching, reshaping the pedagogical approaches and administrative roles in learning, teaching and testing internationally as well as in Turkey. This unprecedented and rapid shift entailed new urgencies and requirements as well as new strategies and experiences for institutions and stakeholders. Clearly, remote testing has become even more visible and critical in higher education institutions in terms of the adaptation of rules and regulations into online platforms, and its presentation, implementation and follow-up stages for multiple stakeholders, i.e. test designers, test-takers, administrators, and teachers during the COVID-19 term. From a sociocultural perspective, individual activities are embedded in and emerge out of the cultural, institutional and historical situations in which they occur. Hence, recognizing how these individual and local testing practices are shaped by and shape the pandemic-driven activity system creates a point of departure for future practices.

Despite the growing number of studies (Clark et al., 2021; Isbell & Kremmel, 2020) in the field, real stories about testing teams' experiences about how high-stakes tests were adapted in response to the unstable and unprecedented COVID-19 disruptions in different

✉ Müzeyyen Nazlı Güngör  
nazlidemirbas@gazi.edu.tr

Mustafa Akın Güngör  
magungor@gazi.edu.tr

<sup>1</sup> Gazi Faculty of Education, Foreign Languages Education, English Language Teaching Programme, Gazi University, Ankara, Turkey

<sup>2</sup> College of Foreign Languages, Gazi University, Ankara, Turkey

institutions are relatively limited (Ockey, 2021). As also rightly argued by Fan and Jin (2020), a clear understanding of the current practices and real stories enable us to take effective remedial steps. Language testing has always been driven by social and contextual factors (Fan & Jin, 2020). For example, Clark et al. (2021) elucidate how decisions were taken considering the test construct, timelines, platforms, remote proctoring, and the role of empirical research in International English Language Testing System (IELTS) test during the pandemic. They highlighted the social and contextual factors, i.e. intense collaboration between testing partners, complexity of at-home testing, and challenges in converting IELTS from delivery at test-centers to an at-home test. Therefore, it is invaluable to explore real pandemic stories and experiences of testing teams as language testing field can benefit from how high-stakes tests are rapidly adapted to online system in local contexts.

Given the changing roles, expanded regulations, and implementation ways during the pandemic, the use of activity theory (AT) to interpret English as a foreign language (EFL) test-designer's experiences in a remote high-stakes test implementation in relation to the community, new roles and rules, mediational artifacts, outcomes, and division of labor in COVID-19 times is a new endeavour. Despite the extensive use and prominent roles of L2 proficiency tests at universities, few studies so far extensively explored how these tests are prepared, implemented, and monitored at universities during the COVID-19 pandemic (Green & Lung, 2021; Wagner & Krylova, 2021). To address this gap, this narrative inquiry seeks to understand how a Turkish EFL test-designer experiences the online L2 proficiency test implementation in unprecedented times from AT framework. Significantly, results reveal three specific activity systems for similar developing countries to understand how online high-stakes testing can be sustained under limited technical and economic conditions beyond pandemic times.

## Literature Review

### High-Stakes L2 Proficiency Tests in Pandemic Times

L2 proficiency tests have the gate-keeping role for test-takers internationally as well as in Turkey. Among the 207 private and state universities in Turkey, 141 of them have College of Foreign Languages (CoFLs) (Higher Education Information Management System, 2020). Organized and implemented generally by these CoFLs, proficiency tests have been used as an indicator of EFL students' language ability to undertake academic work in their fields. Each

university is autonomous to determine this norm through institutional rules and regulations for their newly registered students (TEPAV, 2015). Alternatively, these students who get an equal score from international L2 proficiency tests, i.e. Test of English as a Foreign Language (TOEFL) or IELTS are also exempted from the test and allowed to continue their education in their departments. The others take one-year intensive L2 education in CoFLs. Hence, high-stakes L2 proficiency tests represent a macro-structure that has a profound impact on the content and instructional practices of tertiary EFL learners (Johnson, 2009).

In light of the pandemic precautions and institutional regulations by Council of Higher Education (CoHE, 2020), CoFLs have reported ad hoc solutions for administering L2 proficiency tests remotely. While some have suspended these tests, others have made considerable concessions in test administration in various aspects. First, the process of designing effective tests has become complex regarding the good test construction governed by institutions' pandemic rules and regulations, item design and construction, scoring criteria, and national ethical standards. For example, some of the CoFLs have strengthened their test offices, admitting experienced EFL instructors with good crisis management skills. In this way, it becomes their responsibility to ensure the quality, reliability, validity, practicality, and authenticity of the tests. Second, the implementation of proficiency tests has demanded a broad perspective behind the solely academic concerns. To illustrate, while determining the implementation stage, technology tools that students will need to use in the test, sociocultural dynamics emerging from the test content and administration, the number and responsibilities of stakeholders, internal and existential conditions, and ethical concerns have to be carefully planned and determined through stringent institutional regulations. Third, in the follow-up stage, EFL teachers are to consider the washback effect of the test and legal issues. All in all, understanding the testing experiences of EFL test designers in various roles in global lockdown will help raise the quality of the online test, critically inform future test administration practice, feedback the effectiveness of testing component in teacher education programs, and sustain the success of L2 testing policy beyond the pandemic term.

L2 proficiency test implementation has been under investigation in several aspects (security, validity and reliability of the test, computer literacy problems, ethical rules, privacy) before and during the pandemic. Fan and Jin (2020) examined how L2 test was developed and implemented in relation to social and contextual factors in China. Challenges such as time pressure in marking and communicating test results, lack of support from the administration, difficulty in building teams, cheating students, and stakeholders' lack of language assessment literacy were

detected in their study. Unsurprisingly, Green and Lung (2021), in their L2 test adaptation stories at BYU-Hawaii in the pandemic, pointed to the technological challenges of test-takers due to their first online test experience and limited technology literacy. To avoid these factors from risking test-takers' score interpretations and placement decisions, several make-up test sessions were allowed depending on the validity of the reason. Among the lessons they take, Zoom will be used for live tests to monitor test-taker's behavior, instructional videos and practice tests to familiarize students with the tools, and more testing staff will be trained to remove the unreasonable burden on one person. Unlike these studies, Wagner and Krylova (2021) reported no technology problem in their new test designed for international teaching assistants for Temple University. Since there was a real interaction between the human interlocutor and the test-taker, test-takers were more relaxed in the speaking part. The reliability of the scoring remained the same and, since placement decisions can be made in advance, the new test was found to be time- and resource- convenient for the university. Nevertheless, deciding on the logistics and procedures such as integrating the software and communication platforms, creating a clear and concise test guideline was the risk factor in the test development phase. Importantly, security issue became the main concern in some transitioned L2 tests (Purpura et al, 2021). As a solution, they took remote proctoring and artificial intelligence measures such as Honorlock for monitoring the actions of the test-taker during the test. For authentication and orientation purposes, these test-takers were required to watch a familiarization video and take a short quiz before the test.

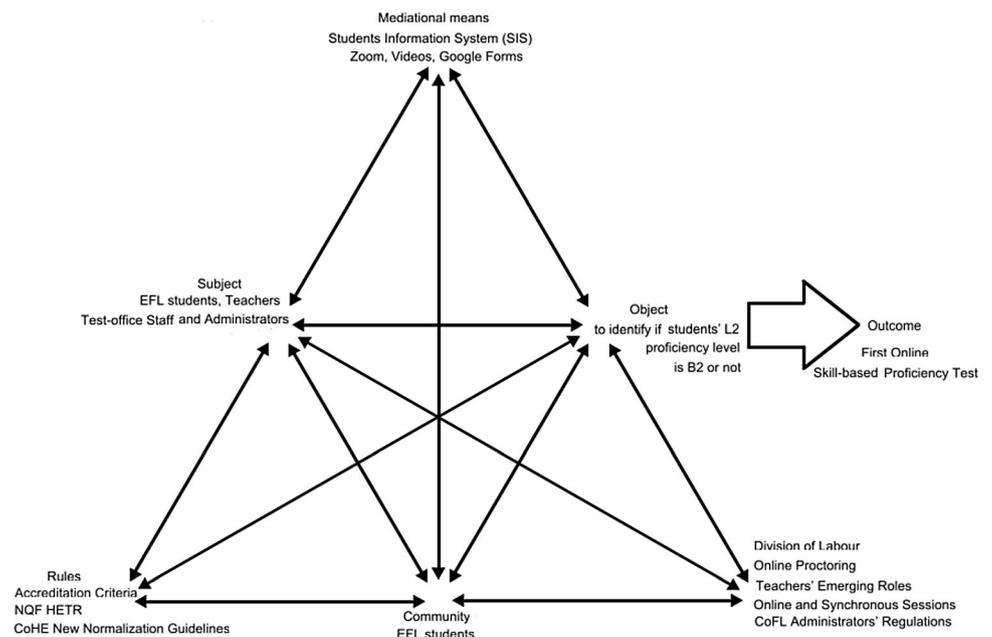
### Activity Theory

Understanding EFL teachers' implementation of high-stakes tests under COVID-19 precautions requires understanding the complex system and contradictions that may drive development in the process. The roles that teachers, institutions, and artifacts play in practice, the conceptual framework of the rules, the mediating tools used to implement the tests, and emerging sociocultural dynamics in the planning and implementation stages constitute an activity system (Fig. 1).

Third-generation AT functions as a powerful and convenient conceptual tool that helps us capture how each component in the complex system affects the other while simultaneously considering the situated activity system as a whole (Johnson, 2009). It enables us uncover inner contradictions between individual actions and the whole activity system. These contradictions may help us better explore the digital transformation of an L2 proficiency test with constraints and affordances to a secure, reliable and valid test administration in unprecedented times, analyze the development of the emerging system based on knot-working sequence, and take a proactive perspective for new learning and developments beyond COVID-19. This study therefore draws on the third-generation AT (Engestrom et al., 1999).

In an attempt to produce ad hoc solutions for the continuity of high-stakes tests in socioculturally and socioeconomically unequal situations due to the outbreak of COVID-19, EFL teachers' actions occur through unstable, constantly negotiated, and transformed relationships of some factors. As shown in Fig. 1, the subject (i.e. EFL students, teachers,

**Fig. 1** Test-designer's activity system



test-office staff and administrators in CoFLs) refers to an individual or group whose agency is the focus of the analysis. Mediational means (i.e. students information system, Zoom, videos, Google Forms, Google Drive) consist of tools which help the object be transformed into outcomes. The object is the orientation of the activity, which will be transformed into the motive for the result. In this study, the object is to identify whether students' L2 proficiency level is B2 or not, grounded in the National Qualifications Framework for Higher Education in Turkey (NQF HETR). The community (i.e. newly registered EFL students at a large state university) consists of the participants who have the same object. Rules (i.e. rules determined by CoFLs parallel to international accreditation criteria and framed by CoHE guidelines (2020) in the new normalization process in global crisis) are those that arise from a local set of social-material conditions. The division of labor (i.e. EFL teachers' emerging roles for being a proctor, head of the test sessions, test-designer, and for preparing synchronous, skill-based proficiency test in pandemic, which differ from face-to-face) refer to the horizontal actions and interactions among the members of the community and the vertical division of power and status (i.e. the CoFL administrators' power that regulate the online test).

As a pedagogical and analytical response to the changing landscape of language testing during the COVID-19, this study aims to understand the implementation of an online high-stakes L2 proficiency test in a narrative inquiry of a Turkish EFL test-designer from AT lens. It also intends to explore the emerging contradictions that may shape effective online test design and administration in pandemic times and beyond.

## Methodology

The exploration of the phenomena in this study was conducted through a narrative inquiry design as Creswell (2012) describes:

... a narrative typically focuses on studying a single person, gathering data through the collection of stories, reporting individual experiences, and discussing the meaning of those experiences for the individual (p. 502).

In this study, the individual stories of a test-designer, Eda (pseudonym), were collected through two sets of data: formal semi-structured interviews and personal communications with her. In formal semi-structured interviews, Eda narrated every detail of the test design and implementation process as well as her beliefs and views, and important lessons learnt from the whole process. Due to the considerable test design and implementation experiences of Eda, this interview lasted a week. In personal communications

subsequent to the interviews, the aim was to clarify her answers to the interview questions, to discuss the contradictions and resolutions in detail, and to develop a broad perspective of understanding the phenomena under exploration. They were all conducted in Turkish to enable Eda to express herself clearly. They were first transcribed verbatim and then translated into English. To ensure the meaning was kept in translations, first the English version was compared with the original one. Then, Eda was required to check the meaning and structure of her answers, and give consent for this study.

All her experiences were restoried based on the three-dimensional space narrative structure elements: *interaction*, *continuity*, and *situation* by Clandinin and Connelly (2000). Subsequently, restories of Eda's experiences were introduced to explore the contradictions and challenges she lived as she was trying to balance roles among her colleagues while concurrently accommodating for expectations placed on her by the administrators for the test in the COVID-19 period. In interaction, personal and social factors in the test implementation were handled regarding her beliefs and views, and existential conditions with other stakeholders' views and purposes. In continuity, a consideration of the roles and responsibilities in previous L2 proficiency tests was related to existing conditions through lessons learnt from those experiences. Next, this synthesis was used as a means for future implementations of the test. In situation, the information was provided with the changing landscape of test rules and regulations, and the heavy burden it caused for Eda. Finally, AT was incorporated as an analytical lens to interpret these results.

## CoFL Context

Eda's stories were set in the context of CoFL of a large state university in the central Anatolia region of Turkey. The institution has offered one-year preparatory L2 education since 1996. The testing content mainly relied on vocabulary and grammar items, and reading and writing skills then. With the employment of the Common European Framework of References for Languages (CEFR) in 2004, the auidial skills (listening and speaking) were integrated. Speaking skill was not tested until 2018 due to the common prejudice against the practicality of the speaking test. September 2018 was the first time when the integration of all four skills was provided in the L2 test. The test system underwent serious changes, with its shift from mechanic to more integrated and skill-based testing. In December 2019, the institution was accredited internationally by Pearson Assured. Aligned with accreditation requirements, the institution is visited for sustainability each year. The testing system which consists of both summative and formative tests is one of the focused components of the visit. Hence, even in unprecedented times, it

was a must for the institution to maintain the accredited level of the L2 proficiency test. It is within this context that Eda's narrations and experiences were situated.

### Eda's Stories: An EFL Teacher and a Test-Designer

Eda was born in a small city in eastern Turkey in late 1980s. Due to her father's job and her placement in a prestigious high school, they had to move to western Turkey. As a result of the national university entrance exam, she was enrolled in an English Language Teaching program of a large state university in middle Anatolia. Following her dreams, she decided to pursue a career as an academic in the city, so she applied for the position of instructor to the CoFL of a large state university upon her graduation with an outstanding academic record.

After she secured her position in 2010, she gained a reputation as a hardworking, successful, and responsible instructor there. Her administrators and colleagues considered her reliable, friendly, and helpful. Hence, her popularity was achieved not only by being strict and critical, but also by being attentive and collaborative. The institution offered a 50-h-long in-service training on "testing and evaluation" to the instructors who were promising in that field, including her. Upon this training, she was assigned to work at the test-office as a test-designer of different language levels. Three years later, she became the head of test-office. She was responsible for the organization, implementation, and evaluation of high-stakes international proficiency test, i.e. TOEFL, IELTS. She participated in various in-service trainings by the Ministry of National Education. Working as the head of test-office for years, Eda has been responsible for high-stakes and low-stakes tests of all language levels in the institution.

In mid-March 2020 when the COVID-19 broke out in Turkey, a burden of responsibility fell on Eda as simultaneously being the test-designer, head of test-office, and EFL teacher. First, she had to plan, implement, and follow low-stakes tests in the emergency remote testing term, and then L2 proficiency test under the "new normal" conditions. Eda's stories below address her testing experiences in an unprecedented term.

## Findings

### Interaction

*"Trust, Cooperation and Communication": The Keys to Crisis Management*

The L2 proficiency test rules and regulations in the pandemic term were shaped by emerging needs and contextual

dynamics of students and teachers due to the shift from face-to-face to remote testing. While the skill-based test had to be adapted to online platform in ways that would adhere to national lockdown policies without compromising the security and validity, there were serious concerns about the sociocultural and socioeconomic realities, i.e. technological resources, economic and health considerations. So, a brainstorming was made among the test designers and administrators, considering the number of test-takers and teachers (administrators, test-givers, test-designer, test-coordinators), aims of the test, the curriculum, and technological affordances and limitations of the institution. Parallel to these realities, Eda, as the head of the test-office and test-designer, contacted several private national companies to conduct the test online. After 3-week-long several piloting studies, the technical and technological limitations of these companies became clearer. For instance, some lacked proctoring feature in the test while some could not allow the higher number of active test-takers, i.e. over 1000 test-takers concurrently. Besides, the timing of the announcement of the university entrance test results and of the new online test rules for the institution clustered, leaving the CoFL less than a month to get prepared for all the workload. The first contradiction arose at this point. As a resolution, she suggested using the limited technological affordances for new purposes and preparing their own online test system.

We were responsible for designing an online test for 1481 test-takers. I hesitantly suggested using the institution's technological infrastructure. We (administrators, coordinators, and me as test-designer) did a SWOT analysis. It appeared we needed 6 Zoom sessions, 50 breakout rooms, and 100 EFL teachers, half of which is heads of the test sessions and the other half is proctors. Although they knew that my suggestion would bring extra workload and risk on all of us, they attached utmost importance to every possibility of my suggestion. Luckily, we had no complexity over each other because of the academic title or position differences. We placed trust in each other. Our administrators were friendly and open to coordination through WhatsApp or Zoom day and night. Consequently, the team, with one vice-director, test-designer (me), and three more coordinators, was assigned to coordinate the whole process. To me, the most important component was this coordination in the preparation stage.

As the story showed, Eda considered the approval as a sign of trust in herself. The first contradiction was resolved based on mutual trust and communication. Second, a detailed planning was made regarding students'

sociocultural and socioeconomic realities, and convenient use of the system both for students and teachers.

### Continuity

*From “But, it’s not My Job”: Changing Roles and Duties to “Yes, We Can!”: Stories of Success to Live By*

Technological tools such as video, Zoom, student information system (SIS), and Google Forms were used for new purposes in the online system. To exemplify, the video was preferred for three main purposes: to increase the test security, to control the synchronous sessions, and to reduce the failures caused by human factor. Test guidelines and rules were also embedded in these videos. However, the fact that there was not a compact test system, which provided secure proctoring and was aligned with national ethical rules, caused the second contradiction for Eda. Knowing that the university already had professional Zoom accounts, she decided to use them as the online test platform where test items were synchronously shared with students. This synchronous sharing, which turned out to be the third contradiction, was vital to provide equal and fair conditions for students (starting and ending the test at the same time, accessing each question simultaneously) during the test. By doing so, she believed the institution would avoid wasting extra money and that test security would be provided via Zoom video-based proctoring. However, computer literacy of the test-takers caused an obstacle. She got into a panic, which appeared to be the fourth contradiction.

Test items were shared with students via videos through “screen-sharing” facility on Zoom. I prepared these videos by embedding a countdown program on the PowerPoint slide. Through a screencast program, the duration was integrated into each slide. And, I saved the screen. Then, I transformed these screencasts into a video and shared it with the heads of the session and proctors on Google Drive. Assuming that students typed their information correctly in the SIS at the beginning of the term, I thought we could use SIS as online answer sheet. But, I was wrong. Two days before the test, some students failing to make configuration in their browsers contacted us for help. After planning every detail, how could I let this ruin the test! In a state of panic and disappointment, I shared an online answer sheet form on Google Forms. Besides, some students phoned the test-office, reporting their spelling errors while typing their e-mails in the SIS. This could have prevented them from getting test links on the test day. Trying to

keep calm, I published an online student form on the institution’s webpage.

Resulting from these emerging needs and mediational artifacts, a new role appeared (i.e. the role of the host) different from the face-to-face practice. The first group included administrators, the head of the test-office, and hosts of the Zoom sessions who were to be contacted in any problem, and to coordinate the whole process during the test (i.e. leading test-takers to breakout rooms). The second group included EFL teachers who served as the heads of the test sessions. Their responsibility was to share the test embedded in the video with students in a session. The coordination skills of this group were vital in synchronously organized multi sessions. Any problem was communicated to these heads, and then forwarded to the first group. Clearly, the second group was responsible for all students and proctors in online sessions. The third group included proctors among EFL teachers in CoFL. Their duties were to check students’ identities and to proctor the breakout session. Eda communicated new roles to these teachers under a lot of stress due to their possible complaints.

Some teachers were anxious about the use of technological tools in the test. Even one teacher rejected his new role arguing: “*But it’s not my job to adapt the technological skills into the test. I do not know how to deal with an unexpected crisis in the test because I am not good at technology.*”

Apparently, the fifth contradiction resulted from EFL teachers’ roles shaped by mediational artifacts which were affected by the lack of a compact L2 test system and some teachers’ low computer literacy. Worried, desperate and exhausted, Eda proposed a line of steps as a resolution. For instance, substitute proctors were assigned for the ones who encountered fast and stable Internet connection or screen-sharing problems during the test. Students’ answers were saved automatically so that they could log in again and continue their test in any unstable Internet bandwidth. The ones who could not solve their problems were allowed to take the make-up test implemented a few weeks later, which was added as a new reason in the test regulations. In this way, no student was unjustly treated. Additionally, a WhatsApp group was set for quick communication and crisis management among the staff in the test. Google Forms were also helpful for students who had problems to make necessary configurations in transforming answers to the online answer sheet. Eda considered these resolutions as stories of success since they enabled to meet the expectations of teachers and students qualitatively and quantitatively.

Personally, I witnessed that the workload and risk factors in online test implementation were more than the face-to-face test. Nevertheless, I must say that I prefer this system for its practicality for the test designers, accessibility for the students, and cost-effectiveness for the institution.

## Situation

### *“Knowing Me Knowing You”: The New Landscape Bounded by Pandemic Rules and Regulations*

The COVID-19 pandemic brought about certain crisis measures in the design and implementation of the test although there was no difference in terms of the test content and items. In other words, what was assessed in face-to-face context was still assessable online today. The evaluation process was also the same as in face-to-face setting, regarding the evaluation criteria, scoring rubrics, and the emphasis of each language skill.

The regulations of the test preparation, implementation, and make-up test in the follow-up stage were varied and increased in amount due to the pandemic. First, the law on the protection of the personal data had to be carefully considered. This law restricts the way of access to the personal data on computers. Taking a reasonable step, Eda took signed consent forms of students before the test for ethical and legal reasons. Second, the emerging roles of the EFL teachers were also regulated by the board decision in the CoFL. Third, the application criteria for the make-up test were revised. To that end, a committee was established to decide on the official applications and Eda was appointed as the chair of the committee. All in all, the new landscape placed heavy burden on Eda, enriching her reconceptualization of remote testing with all great details.

## Discussion

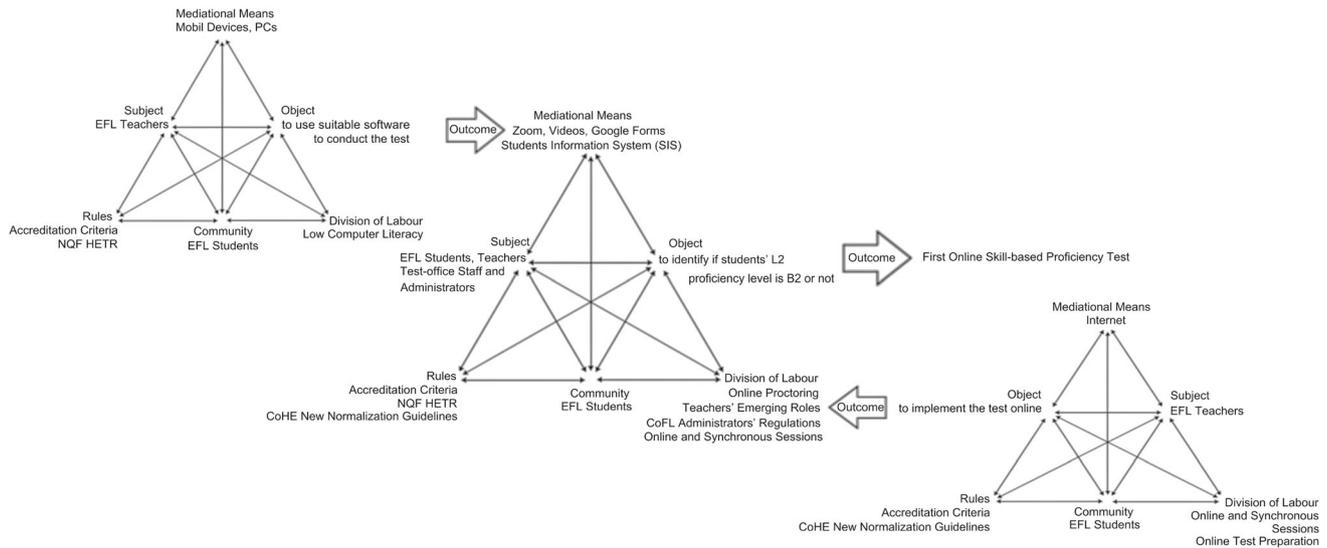
This study aimed to understand the implementation of a high-stakes L2 proficiency test in a narrative inquiry of a Turkish EFL test-designer from AT lens. The other aim was to explore the emerging contradictions that shaped effective remote test design and administration in pandemic times. Overall, Eda’s stories showed how her reconceptualization of remote testing was shaped not only by her previous testing knowledge but also by her colleagues’ intervention and contextual contradictions.

In each contradiction (1st contradiction: the implementation and transformation of the test into online platform; 2nd: the need to use Zoom; 3rd: the necessity for starting synchronous test sessions; 4th: the low level computer

literacy of test-takers; 5th: teachers’ emerging roles and their reactions to these) which was embedded in the three-dimensional space narrative structure, the use of AT helped us identify unstable and unpredictable activities of the subject and increase our awareness of creative adaptation of face-to-face system to remote testing beyond the pandemic.

In interaction, we identified Eda’s tensions in search of alternative solutions in the context of administration and community. For example, the object was to identify whether the level of newly registered students was B2 or not in the online L2 proficiency test. On one hand, Eda had to adhere to both the existing rules, such as aims and content of the test in line with L2 proficiency (B2 level) by NQF HETR and accreditation requirements. On the other hand, she had to plan the whole online test process in a limited time. This contradiction brought serious risks for the administrators and test-office. As Eda’s biography demonstrates, they valued her previous testing knowledge and her personal and professional identity. Hence, they trusted her at all costs and expanded her authority. As her immediate responses to the contradictions and creative use of artifacts proved to be effective in crisis times, her beliefs were positively affected in her own activity system. Similar to the IELTS team (Clark et al., 2021), it was a critical opportunity that increased her awareness of the importance of mutual trust, collaboration and communication in administration in crisis times. This confirms Xu and Liu’s (2009) findings that is the more autonomous teachers are in testing, the more risks they take in their testing practice. This emphasizes the necessity of the test-designer’s agency via dialogic mediation in the design and implementation of innovative approaches.

In continuity, resolutions followed a line of horizontal actions and interaction among the subject and community members. These horizontal actions were related both to the employment of emerging roles across social and economic realities of the community, and for technological ease for the subject and community. AT perspective enabled us to understand how Eda’s reactive approach created mediational artifacts with new purposes for test security. To specify, videos were embedded in PowerPoint slides and Zoom was used for secure proctoring purposes (Wagner & Krylova, 2021). Google Forms was used as an online answer sheet in the last minute due to the computer literacy problems. She thereby provided assisted performance for test-takers’ configuration through online remedial guidance as also employed by Wagner and Krylova (2021). As seen, what made these artifacts effective for these teachers was Eda’s previous interest in technology, creative use of them for problem-solving purposes, and other activity systems emerging in crisis times (Fig. 2). However, these artifacts negatively affected teachers’ approach towards emerging



**Fig. 2** Joint activity systems

roles and exhausted Eda mentally. Johnson (2009) noted that both physical activities and language are important to help us identify how the subject and community understand the contradictions they face in an activity system. Similarly, Eda communicated them for their hesitations and objections to better understand their points, and supported them via technological and technical affordances in collaboration with distance education unit, considering the wellbeing of teachers and students in crisis time. Awareness of the hierarchy culture rooted in Turkish culture stimulated her endeavors to create dialogic mediation with teachers. While many were open to the mediation and interaction, some still rejected to perform these roles. Eda sensed the fragility of social relations and remote testing system, rather than the workload, left her feeling exhausted.

AT analysis enabled us to identify Eda's powerful sense of agency and responsibility in her urgent, short-term, and high-stress problem solving and decision-making under pressure to sustain the continuity of the test. Clearly, each resolution supported the continuity of the next step in the test administration. For us, these were essential, pragmatic, intentional, and creative resolutions adding value to our comprehension of future testing practices. As argued by Isbell and Kremmel (2020), this led her reactive approach to be moulded into proactive one for future remote testing practices. Now the question is: will these resolutions be sustained in the future remote testing systems under the new normal?

In situation, remote testing took place in two main places: at home and on the Internet. Clearly, the board decisions for valid reasons in the revised make-up test regulations and privacy in test administration were affected

by conditions of test-takers and givers in these places. This space proved how the unprecedented nature and essential role of test places contributed to Eda's reconceptualization and her personal practical knowledge. Stories displayed issues of ensuring synchronous and secure test implementation, and the existing technological and technical infrastructure for the continuity of the test had to be concurrently addressed in online setting (vertical division of power and status). As highlighted in other studies (Purpura et al, 2021), this indeed caused security, economic and health hesitations, fairness, and technological ease to outweigh the test validity and reliability, which confirmed the complexity and problematic aspect of remote testing (Clark et al, 2021). Also, the accreditation and NQF HETR requirements were reported to be the sources of this complexity.

## Conclusion and Implications

This study is important for a developing EFL country to prove remote high-stakes testing could be achieved in a large state university with its own infrastructure in unprecedented times. Considered the lack of quick access to the massive technical infrastructure, and sociocultural and socioeconomic inequalities of the subject and community, it was a challenging but successful and informative experience. The activity system of the test-designer was influenced by two other activity systems (Fig. 2). All their elements can be a model both for the replication of further studies and successful implementation of remote high-stakes test in other developing countries.

Resolutions imply important lessons to be taken in this study at macro and micro levels. In the first one, macro remote testing policies such as ensuring fair and economic access to the tests, and strengthening the test offices qualitatively and quantitatively, should be prepared for future disruptions. In the latter, administrators should provide the test-coordinators with autonomy and dynamism in crisis times (Clark et al., 2021). Effective communication skills among the subject and community members, and quick decision-making skills of test designers are also proved to be critical in overcoming teachers' psychological and students' technical barriers. Moreover, continuing support (technology integration into testing) should be offered to teachers to help them accommodate to new roles in unprecedented situations.

New contagious diseases are foreseen (Isbell & Kremmel, 2020). Hence, future studies should examine how test security and validity are warranted in online high-stakes test, how online test constructs are influenced by multiple perspectives on language knowledge and use, and how digital transformation in high-stakes tests shapes the test task design. Additionally, the washback effect of this transformed test can be examined in a new study as for how it shapes the instructional practices in distance education and influences teachers' perceptions of remote L2 proficiency tests. Data from these studies will inform the future of language testing field.

In conclusion, AT perspective enabled us to see how a test-designer's individual activities were interwoven and produced "knotworking" in the local context (Fig. 2). Four activity systems in this context may be influenced by other contexts. Therefore, more AT studies are needed in testing to explore new contradictions, emerging roles and needs, and possible resolutions based on stakeholders' experiences in crisis times. This contributes to our understanding of "knotworking" in multiple AT systems.

## Limitations

We acknowledge a few limitations in this study. First, we only focused on the activity system of the test-designer due to practicality. Future studies may explore the activity systems of different stakeholders, i.e. administrators, test-givers and takers. Second, proficiency tests practice may vary in different universities. Hence, caution is required when interpreting findings of this study. Third, data triangulation were provided through two sources: narrations and personal communication. Complementary data source may be added in future studies.

**Acknowledgements** We would like to thank the special issue guest editors: Shulin Yu and Hao Xu, and the anonymous reviewers for

their feedback. We appreciate the hard work Eda and her team did to administer the L2 proficiency test remotely despite limitations in times of COVID-19.

## References

- Clandinin, D. J., & Connelly, F. M. (2000). *Narrative inquiry: Experience and story in qualitative research*. Jossey-Bass.
- Clark, T., Spiby, R., & Taviri, R. (2021). Crisis, collaboration, recovery: IELTS and COVID-19. *Language Assessment Quarterly*, 18(1), 17–25. <https://doi.org/10.1080/15434303.2020.1866575>
- Council of Higher Education [Yükseköğretim Kurulu]. (2020). *New normalisation period in pandemic [Kuresel salgında yeni normalleşme süreci]*. Retrieved November 20, 2020 from <https://www.yok.gov.tr/Documents/Yayinlar/Yayinlarimiz/2020/kuresel-salginda-yeni-normallesme-sureci-2020.pdf>
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.
- Engestrom, Y., Engestrom, R., & Vahaaho, T. (1999). When the center does not hold: The importance of knotworking. In S. Clandinin, M. Hedegaard, & U. J. Jensen (Eds.), *Activity theory and social practice: Cultural-historical approaches*. Aarhus: Aarhus University Press.
- Fan, J., & Jin, Y. (2020). Standards for language assessment: Demystifying university-level English placement testing in China. *Asia Pacific Journal of Education*, 40(3), 386–400.
- Green, B. A., & Lung, Y. S. M. (2021). English language placement testing at BYU-Hawaii in the time of COVID-19. *Language Assessment Quarterly*, 18(1), 6–11. <https://doi.org/10.1080/15434303.2020.1863966>
- Higher Education Information Management System [Yükseköğretim Bilgi Yönetim Sistemi]. (2020, November 20). Retrieved from <https://istatistik.yok.gov.tr/>
- Isbell, D. R., & Kremmel, B. (2020). Test review: Current options in at-home language proficiency tests for making high-stakes decisions. *Language Testing*, 37(4), 600–619.
- Johnson, K. E. (2009). *Second language teacher education: A sociocultural perspective*. Routledge.
- Ockey, G. J. (2021). An overview of COVID-19's impact on English language university admissions and placement tests. *Language Assessment Quarterly*, 18(1), 1–5.
- Purpura, J. E., Davoodifard, M., & Voss, E. (2021). Conversion to remote proctoring of the community English language program online placement exam at Teachers' College, Columbia University. *Language Assessment Quarterly*, 18(1), 42–50.
- TEPAV. (2015). *Türkiye'de yükseköğretim kurumlarındaki İngilizce eğitimi*. British Council. Retrieved October 27, 2020 from [https://www.britishcouncil.org.tr/sites/default/files/20160211\\_english\\_he\\_baseline\\_study\\_final\\_tr.pdf](https://www.britishcouncil.org.tr/sites/default/files/20160211_english_he_baseline_study_final_tr.pdf)
- Wagner, E., & Krylova, A. (2021). Temple University's ITA placement test in times of COVID-19. *Language Assessment Quarterly*, 18(1), 12–16.
- Xu, Y., & Liu, Y. (2009). Teacher testing knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, 43(3), 493–513.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.