



## *Doing language testing: learner-initiated side sequences in a technology-mediated language learning environment*

Simin Ren & Paul Seedhouse

To cite this article: Simin Ren & Paul Seedhouse (20 Feb 2024): *Doing language testing: learner-initiated side sequences in a technology-mediated language learning environment*, Classroom Discourse, DOI: [10.1080/19463014.2024.2305446](https://doi.org/10.1080/19463014.2024.2305446)

To link to this article: <https://doi.org/10.1080/19463014.2024.2305446>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 20 Feb 2024.



Submit your article to this journal [↗](#)



Article views: 210



View related articles [↗](#)



View Crossmark data [↗](#)

# Doing language testing: learner-initiated side sequences in a technology-mediated language learning environment

Simin Ren  and Paul Seedhouse 

School of Education, Communication and Language Sciences, Newcastle University, Newcastle upon Tyne, UK

## ABSTRACT

We present *doing language testing* sequences; L2 learners decide for themselves to test themselves or each other explicitly on new linguistic items, outside the official task cycle with no professional present, investing extra time and energy. We examine how and why pairs of learners do this, and its impact on their learning. They use an App to learn Chinese language and culture whilst cooking in their university dormitory kitchens, receiving multimedia instructions and help from a tablet. Using a mixed-methods research design, we asked: How is the practice of *doing language testing* organised in interactional terms? Using multimodal CA we found: learners organised the interaction themselves and introduced their own learning interests. Some self-tested and some peer-tested; some consulted the system and some did not. To ascertain the reasons why the learners decided to *do language testing*, we used post-hoc interviews and found their major motivation was to have an improved learning experience. We then compared their vocabulary post-test score gains with those who did not *do language testing* and found they made significantly higher gains with a learning advantage. The study shows that some students are motivated to *do language testing* in order to enhance their task performance and learning experience, in which they succeed.

## ARTICLE HISTORY

Received 8 February 2023  
Accepted 9 January 2024

## KEYWORDS

Doing language testing; peer assessment; technology-mediated language learning; L2 learning; language testing

## 1. Introduction

This article introduces and studies sequences which have not previously been discussed in the Conversation Analysis (CA) literature, namely '*doing language testing*' sequences. When *doing language testing*, L2 learners in pairs or groups elect to test themselves or each other explicitly on newly-encountered linguistic items, when this is not part of the envisaged pedagogical design or task structure; they decide for themselves to devote extra time and energy to this. In this setting, their use of '*doing language testing*' was entirely unexpected. In CA terms, these are side sequences (Jefferson 1972), while from

**CONTACT** Paul Seedhouse  paul.seedhouse@ncl.ac.uk  School of Education, Communication and Language Sciences, Newcastle University, Room 3.01, Newcastle upon Tyne NE1 7RU, UK

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

the TBLT perspective they can be regarded as ‘incidental focus on form’ (Ellis 2001), where language learners’ primary focus is on forms and there is a distribution of attention to specific linguistic forms. We analyse the organisation of the interaction using multimodal CA. We also examine why the learners choose to do this using post-hoc interviews and check whether they may be gaining a learning advantage by comparing their post-test score gains with those of the majority of the cohort, who did not *do language testing*.

The pedagogical design of the Linguacuisine/ENACT software used in this study <https://enacteuropa.com/> was based on the principles of Task-Based Language Learning and Teaching (TBLT) (Ellis 2003; Long 2015). TBLT tasks have previously been employed for online/computer assisted learning in a number of ways (Thomas and Reinders 2010; González-Lloret and Ortega 2014). In our study, however, principles of TBLT are not employed online or in the classroom; we use the ‘real-world digital environment’ of the kitchen (Seedhouse et al. 2019) in which learners engage all multimodal resources, including all five senses, in the cooking task. Linguacuisine/ENACT app materials have been produced for learning a range of languages; in this project materials were created for learning Chinese language, culture and cuisine. Therefore, we refer to this as the Chinese Digital Kitchen (CDK). In order to clarify the context within which *doing language testing* occurs, we briefly sketch the technology-assisted language learning environment in section 5.

At the time the learners *do language testing*, they have already taken a pre-test and pre-task (which introduces them to the L2 names of ingredients and utensils) and are aware they will do a post-test after the task cycle (see Figure 1 for details). They are therefore in a position to proceed to the main cooking task, and the majority of learners (83.3%) elect to do so. However, a substantial minority (16.7% or 12 learners) elected to do language testing at this point. Why do they do so? We had two hypotheses. It can be hypothesised that they decide to undertake the *doing language testing* side sequences

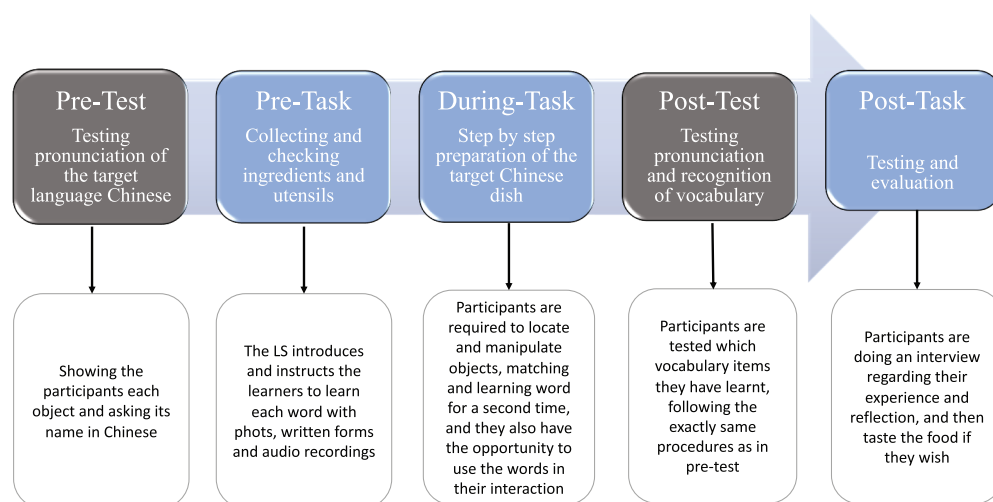


Figure 1. Task and test cycle in the Chinese Digital Kitchen (CDK).

as preparation for the post-test, possibly with the motivation of improving their performances and/or scores. Since we had the pre-test and post-test data, we compared the test scores of the sub-cohort of learners who had undertaken *doing language testing* side sequences with the sub-cohort which had not and asked if attainment correlates in any way to the additional sequences. A competing hypothesis was that they *did language testing* in order to be sure they could locate the required equipment and ingredients accurately during the cooking task in order to complete it more efficiently and have an improved learning experience. It is also possible that learners had both motivations.

## 2. Literature review

### 2.1. CA studies of language testing

The study<sup>1</sup> touches on a number of areas in the worlds of language learning, teaching and testing. Whereas there have been many CA studies of language learning in informal (or technology-mediated) settings involving no teachers, our search did not reveal any CA studies involving language testing with no professional present.

Conventionally, *testing* refers to the process of measuring learners' knowledge, skills, or abilities in a specific domain with clear criteria, which involves designing and administering tests or exams to evaluate learners' performance, and aims to assess various language skills including speaking, writing, reading, and listening (Bachman 1989; Fulcher and Davidson 2007). *Assessment*, on the other hand, encompasses a broader and more comprehensive view than testing. It involves multiple sources of evidence, such as classroom observations, portfolios, interviews, projects, and self-assessments, and it aims to provide a comprehensive and ongoing evaluation of learners' language development, allowing educators to understand learners' strengths and weaknesses, and providing targeted feedback (Brown and Abeywickrama 2010). A test is therefore seen more as a 'product' and an assessment is seen as a 'procedure'. In relation to the context of this study, Bachman and Palmer (1989) also add that language assessment takes place in a real-world that is often unpredictable and in which conditions and individuals who are involved in or affected by the assessment change over time. We therefore contribute to the investigation of how L2 learners configure learning opportunities themselves without the guidance or presence of teachers or professionals.

CA studies of language testing have been of interview tests or oral proficiency interviews, conducted in L2 between a professional interlocutor or examiner and candidate(s), for example, Kormos (1999) focused on simulating conversations in oral-proficiency assessment of role plays and non-scripted interviews in language exams, Kasper and Ross (2007) focused on interaction troubles ensuing from interview candidates misconstruing or mishearing questions in oral proficiency interviews; Seedhouse and Nakatsuhara (2018) provide a unique dual perspective on the evaluation of spoken discourse in that it combines a detailed portrayal of the design of a face-to-face speaking test with its actual implementation in interactional terms, using empirical extracts of interaction from authentic IELTS Speaking Tests. They also illustrate how the interaction is organised in relation to the institution aim of ensuring valid assessment. Youn and Burch (2020) discussed how to use CA to expand the epistemologies and validity evidence

of language assessment, as they attempt to combine an inherently *emic* field (CA) with an inherently *etic* field such as language testing and assessment; this could therefore address a number of the epistemological and methodological challenges.

More importantly, there was nobody telling the participants to conduct such sequences, and there was no mark or certificate given to them unlike a standard test. Therefore, due to the lack of the existing literature, the present study defines 'doing language testing' within TBLT as: learners nominate either themselves or each other for explicitly testing their linguistic knowledge in terms of the target words, when this is not required by the task structure. Language testing does not include this phenomenon, which has not yet been studied in language testing, language acquisition, or relevant CA studies.

## 2.2. Peer-assessment

Peer-assessment (PA) has been attracting considerable attention for more than four decades in both education and applied linguistics (Chang and Lin 2020; Hoffman 2019; Meletiadou and Tzagari 2016; Topping 2018) due to the increased focus on and popularity of learner autonomy and independence (Peng 2010). As an alternative form of 'assessment for learning' promoting learner-centred assessment, PA is a process in which a group of individuals test their peers, which may or may not involve agreed criteria between teachers and students (Falchikov 1995). Particularly, it has been regarded as a planned educational arrangement where students judge a peer's performance quantitatively, such as providing a peer with grades or scores, and/or qualitatively, for example, by providing a peer with oral or written feedback (Topping 2017). PA is argued to have significant pedagogical value because it enables learners to take part in assessment by evaluating their own and also their peers' learning process and products (Bryan and Clegg 2019; Patri 2002), stimulating students' motivation and promoting critical assessment skills (e.g. Freeman 1995; Pope, 2001; Topping 2018), increasing effective self-regulation to identify mistakes and develop strategies to address them (Zamora, Suárez, and Ardura 2018), etc. However, there are also some critiques of PA such as: they are not regarded as experts and the accuracy of PA varies (Reinholz 2016); it requires continuous and repeated practice for learners to become competent peer assessors (Andrade 2016), whereas peers do not have the kind of authority and subject knowledge that teachers have (Topping 2017) leading to their feelings of discomfortable or insecure (Cheng and Warren 2005; Orsmond and Merry 1996). Therefore, as Topping (2009) and Patri (2002) have argued, PA should not be a substitute for traditional teacher assessment but may add value as a supplement to the learning and assessment process, especially since its effective use seems to depend on a wide range of factors including students' attitudes, language levels, a critical thinking skills (Peng 2010).

In a word, the primary purpose of peer assessment is to provide feedback, encourage self-regulation, promote critical thinking, and foster collaborative learning among individuals within a group. By contrast, the *doing language testing* investigated in this article and invented by the participants themselves aims to produce an improved learning experience and performance. To summarise, there is a substantial literature on teacher-organised PA in the fields of education and applied linguistics which provides evidence of its relative advantages and

disadvantages. However, we have been unable to locate any studies of PA practice which is initiated and organised by the learners themselves spontaneously outside of the official task or curriculum; this is what is involved in *doing language testing*. Another gap is in studies looking at whether PA of L2 speaking results in enhanced performance, although Meletiadou (2021) suggests moderate gains associated with PA in the writing performance of 200 Greek Cypriot EFL students. Published work has disproportionately targeted learning of L2 English, with only a handful of studies focusing on other languages.

### 2.3. *Speech exchange system*

As the nature of cooking tasks in a real-world kitchen and the features of the digital Linguacuisine app have had an impact on the nature of participants' interaction, and such impact can be seen in the ways in which participants designed their turns. Hence, our analysis follows Sacks, Schegloff, and Jefferson (1974, 729–731)'s definition and description of speech exchange system, which is:

'the organisation of practice addressed to these issues – turn organisation and turn-taking organisation, sequence organisation of repair, the organisation of word selection, overall structured organisation, and others, in the options which they shape and the practices they make available – constitute a spate of interaction recognisable as "conversation", as "interview", as "meeting", as "giving a speech", as "interrogation", etc. Termed as "speech exchange system" (SES henceforth) and as particular, here-and-now-with-these participants instances of these'. In general, SES investigates the turn-taking organisation on the sequential level, seeks to uncover the ways in which an L2 speaker's talk is demonstrably oriented to that speaker and/or their interlocutor(s), where interlocutors predominantly speak one at a time and establish connections during their respective turns (Sacks, Schegloff, and Jefferson 1974) as well as its impact on the trajectory of the interaction (Gardner and Wagner 2004). Sawchuk (2003) also emphasised that the analysis of a speech-exchange system offers a powerful way of taking a close look at and generating warrantable claims about the nature of learning as a distinct social phenomenon.

Furthermore, the configuration of an SES can be influenced by various institutional goals. Within the pedagogical design and framework (TBLT), which outlines the fundamental SES within the digital kitchen environment and without professionals or teachers present, specifically the Linguacuisine app, the initial turn involves verbalising cooking instruction and each pair is then expected to carry out pertinent subsequent verbal and/or non-verbal actions. However, the activity of conducting *doing language testing* is actually a spontaneously devised endeavour and an optional side sequence to its basic SES. Consequently, we draw comparisons regarding the organisational aspects of the *doing language testing* procedure by the two pairs in this article. In this regard, we aim to investigate how participants, during the execution of such *doing language testing* sequences, construct their individual turns and actions in relation to one another. We also delve into how these sequences are structured within their interaction, as defined by Sacks, Schegloff, and Jefferson (1974). The manner in which participants shape these sequences is influenced by the characteristics of the environment and the nature of the task at hand.

## 2.4. Discussion

We do not review here the relevant literature on educational technology and computer-assisted language learning (CALL) because a) it is of minor relevance to the focus on *doing language testing* b) this is available in Seedhouse (2017) and Seedhouse et al. (2019). However, we note at this point that the digital technology design approach we use is known as ‘ambient’ ‘pervasive’ or ‘ubiquitous’ (Olivier et al. 2009; Weiser 1999) in that the technology assists people to carry out real-world tasks – in this case cooking in a kitchen. This is highly relevant as the pedagogical framework of the Linguacuisine project is in fact the integration of TBLT and CALL, which has contributed to pedagogical developments in the field of foreign/second language teaching and learning (Thomas and Reinders 2010). Many previous studies on the integration of TBLT and CALL have focused on their use in classroom pedagogy, but there are few studies which explore their use outside a classroom context (Preston et al. 2015; Seedhouse et al. 2013, 2014). Being a real-world digital language learning environment, as Seedhouse (2017: 10) pointed out, the key features are: 1) participants physically carry out real-world tasks (using real-world equipment) underlying TBLT framework; 2) technology is utilised to facilitate performance of the task; 3) the system is pedagogically designed to help participants’ target language learning; 4) participants are supposed to receive target language input from some source and can learn some aspects of the language for their task performance, while in the meantime, they have the opportunity to physically touch and manipulate real-world objects and learn the names of the objects in the target language; 5) a task-cycle should be planned and there should also be a real-world outcome (in this case, cooking a Chinese dish).

A previous study of the Finnish Digital Kitchen proved that the linguistic input of the computer alone does not trigger learners’ orientation to language or language learning (Kurahila and Kotilainen 2020). Despite receiving verbal instructions, learners are in fact immersed in an ecological environment with all the resources and affordances that a cooking activity could offer, which means that learners may also need to solve practical problems in relation to cooking or technological issues. Such negotiations can, in turn, lead to discussions about language (Kurahila and Kotilainen 2020), which in our case, they transitioned to a *doing language testing sequences* that were decided for themselves to test themselves or each other explicitly on new linguistic items.

In our study, participants’ use of *doing language testing* was entirely unexpected which happened after their pre-task and prior to the during-task (see Figure 1). We therefore attempt to fill in the research gaps in the following aspects: 1) CA studies of language testing sequences in the real-world technology-mediated context; 2) peer assessment practice which is initiated and organised by the learners themselves outside of the official task or curriculum, with an underexplored target language Chinese; 3) in real-life technology-mediated context, how participants display and construct their turns within their speech exchange systems; 4) adding to previous Digital Kitchen projects studies, L2 learners’ *doing language testing* as a completely new and original part.

Therefore, we examine how and why pairs of learners do this in the CDK context, and its impact on their learning. We then move to explain how mixed-methods design can help answer our research questions.

### 3. Research design

The phenomenon investigated is the unexpected one of *doing language testing*, in which L2 learners in pairs or groups elect to test each other explicitly on newly-encountered linguistic items, when this is not part of the envisaged pedagogical design or task structure. Having noticed this unexpected phenomenon, we aimed to understand its nature and rationale, therefore, we ask the following research questions:

How is the practice of *doing language testing* organised in interactional terms? To answer this, we analyse the organisation of the interaction using multimodal Conversation Analysis (CA).

Why did these learners decide to *do language testing*? We answer this by using post-hoc interviews.

Did they gain a learning advantage by *doing language testing*? We compare their post-test score gains with those of the majority of the cohort, who did not do language testing. All learners took pre-tests and post-tests on the same 27 vocabulary items.

This study employs a mixed-methods research design to explore the effectiveness of *doing language testing* in aiding L2 learners' understanding and application of newly encountered linguistic items in a real-life technology-mediated environment. The study focuses on learners' independent decisions, without the presence of teachers or professionals, which contains the following elements (Seedhouse 2022):

- (a) A language learning task (cooking a Chinese dish) using the principles of TBLT and including a 3-phase task cycle, which provides the L2 learning input.
- (b) A pre-test/post-test cycle which wraps around the task cycle as shown in [Figure 1](#). This provides the quantitative evidence of the *product* of learning.
- (c) Video and audio recording of the test and task cycle undertaken by the participants. This facilitates CA studies of the *process* of learning through multimodal analyses.
- (d) Interview evidence from the participants about their learning and motivations.

The specific type of mixed-methods design used is 'triangulation design' (Creswell and Plano Clark 2017, 59–79) which can be glossed as follows: this brings qualitative and quantitative methods to bear on a research problem in a single phase in order to better understand it. Typically, the researcher collects the two types of data separately but at the same time, then brings them to bear on the problem, giving each element equal weight. In terms of triangulation, the perspectives which are combined in our approach are: a) the emic perspectives of the participants engaged in language learning talk, established by CA; b) the post-hoc self-report perspectives of the participants via interviews about their activities c) the 'testing perspective' on changes in terms of participants' learning states, established by comparing pre-test vocabulary scores to post-test scores.

It is important to examine the task and test cycle as all learners chose to do language testing at exactly the same point in the cycle. The task cycle is separate from the test cycle, although one is wrapped around the other ([Figure 1](#)).

The task structure consisted of pre-task, main task and post-task. In the pre-task, the system introduces the learners to vocabulary items needed in the main task. Following its introduction in the pre-task, each vocabulary item is then repeated verbally by the system



at least once during the main task (the cooking session) as part of the cooking instructions, thus providing further input. The participants may also produce the vocabulary items when speaking to each other as they conduct the task. The system requires the learners to physically manipulate the objects during the tasks, whilst the task design provides the opportunity (but not the necessity) for participants to employ the vocabulary in their joint dialogue. In the post-task, the participants sample and evaluate the food that they have cooked. This gives them a further opportunity (but not obligation) to employ vocabulary learnt. So each learner hears the name of each vocabulary item a minimum of 2 times from the system, but there is no maximum. Learners can continue asking the system to repeat the name of an object as many times as they choose to assist their learning.

The pre-test comes before the test cycle and establishes in the participants' minds a need to learn the vocabulary items. If they do not know the item, they may realise that they will need to learn the item in the pre-task, and so an information gap and a motivation is created. After the main task comes the immediate post-test, when participants are able to evaluate which vocabulary items they have learnt, following the identical assessment procedures to the pre-test. The *doing language testing* sequences examined in this article are introduced incidentally by the participants themselves after their pre-task but prior to the during-task phase.

#### 4. The research setting and sample

The four learners involved in the present article are L2 learners of Chinese in China; SHASHA is more proficient than YEDA within the intermediate level, while the second pair WEN & LIU are both at advanced level. They are in the kitchen of a student dormitory and their language learning task is to cook a Chinese meal (*Eggplant Stir Fry*) using the equipment and ingredients in the kitchen, as well as learning a related L2 vocabulary set. This is not a virtual online environment. The learners are performing the real-world task of cooking real food in a real kitchen, but it is a digitally-mediated environment in that they must interact with a digital system (the Linguacuisine/ENACT app) in order to complete the task.<sup>2</sup> Seedhouse (2017) calls this a real-world digital environment. They are receiving



Figure 2. View of ENACT app tablet screen with Chinese instructions.

instructions and help on demand by video, audio, photos and text in L2 Chinese from a tablet, as in Figure 2 below.

Figure 2 shows a typical cooking instruction as a screenshot from the tablet which the learners can see. There is a researcher present who is recording the interaction but does not provide any help or feedback of any kind. The learners have not been given any instructions on how to interact with each other or with the tablet but have been shown which buttons can be pressed on the tablet to receive help in terms of video, audio, text and photos in Chinese in relation to the recipe.

As shown in the above Figures 3 and 4, the data were collected in the communal kitchen in the participants' dormitory in a university in China. The cohort consisted of 72 international students of L2 Chinese. Participants were 43 males and 29 females in total, age ranged from 18 to 40 years old, and their exposure to Chinese varied between 2 months and 68 months, with a mean of 13 months (1 year and a month). We tried to pair the participants so

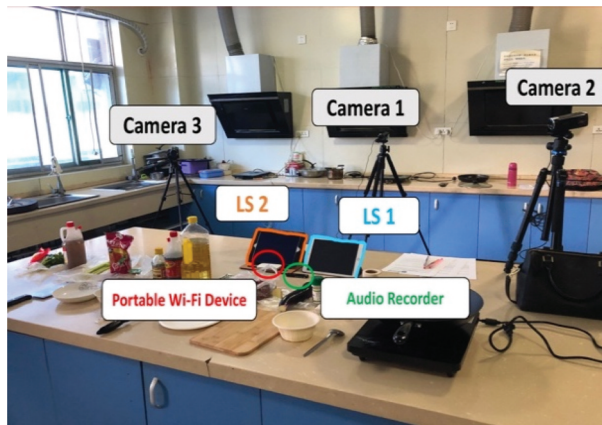


Figure 3. Research setting 1.



Figure 4. Research setting 2.

that one had a higher language proficiency than the other. In most cases, the participants did not have a common L1 and spoke Chinese or English L2 the whole time. Participants who had a common L1 (as in Extract 2) were requested to speak Chinese L2 or English L2, whereas it sometimes happened that they code-switched between Chinese and English.

Of the whole cohort, 16.7% or 12 learners elected to *do language testing*. In the corpus of 6 ‘doing language testing’ sequences by 12 learners (6 pairs), there are variations in how interaction is managed. There are two basic variations: 3 pairs use the system as a point of reference and help and 3 pairs do not. In 4 pairs, one learner acts as ‘examiner’, nominating items for the other learner to name in L2, whereas in 2 pairs one learner at a time self-tests. We examine two extracts. The first features self-testing without reference to the system; the second, one student nominates items for the other, who sometimes gets help from the system.

## 5. Analyses of interactional data

### 5.1. Self-testing sequences

The following pair of participants, SHASHA & YEDA (see Figure 5), two intermediate-level participants, initiate a ‘self-testing sequence’ – for all of the 27 items of target vocabulary. Before commencing their *doing language testing*, SHASHA & YEDA completed the pre-test phase and enhanced their acquaintance with the target word ‘chao cai chan’ (spatula) during the pre-task phase, utilising the Linguacuisine system (LS 2) for its pronunciation (line 024-line 027). Following a brief pause indicated in line 029, YEDA’s utterance indicated the completion of *pre-task* phase, and they are officially supposed to proceed to the *during-task* phase to start cooking and working with LS 1 now. However, they chose not to do so but transitioned into a *doing language testing* sequence, starting in line 030. The reasons for this are explored in post-hoc interviews in Extract 3.

	<b>SHASHA</b>	<b>YEDA</b>
<b>HSK Level</b>	HSK 4 (intermediate)	HSK 3 (intermediate)
<b>Ethnicity</b>	Tajikistani	Kazakhstani
<b>Gender</b>	Female	Female
<b>Age Group</b>	18-25	18-25
<b>University Programme</b>	Cross Cultural Communication (MA)	Cross Cultural Communication (MA)
<b>Arrived in China</b>	09/2017	09/2017
<b>Length of Stay in China</b>	2.5 years	2.5 years

Figure 5. SHASHA & YEDA’s information.

**Extract 1** 'ni shuo shi shenme' (you tell me what it is) Appendix 2

**Extract 1a**



024 shasha (1.4) **SHASHA** clicks on the audio of 'chao cai chan' (spatula) on **LS 2**  
 yeda **YEDA** looks at the **LS 2** with her right hand on it  
 fig FIG 1 relates to line 024  
 fig FIG 2 relates to the system screen



1

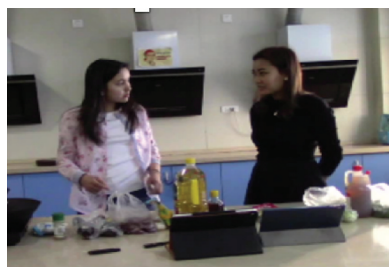


2

025 **LS** *chao:: cai:: chan::*  
 026 **YEDA** *chao cai:: chan*  
 027 **SHASHA** [*chao cai:: chan*]  
 028 (0.3)  
 029 **YEDA** *hao le, wan le.*  
 ok, finish.  
 030 **SHASHA** *shi yi shi, ni neng ji zhu (0.2) ji ge?*  
 just give it a try, see how many you can remember?  
 shasha **SHASHA** points at the objects on right hand side  
 yeda **YEDA** looks at **SHASHA**  
 fig FIG 3 relates to line 030 (shasha's non-verbal action)  
 fig FIG 4 relates to line 030 (yeda's non-verbal action)



3



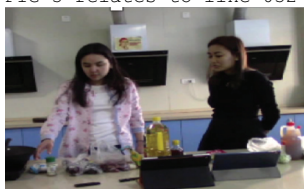
4

031                    (1.0)  
 yeda                **YEDA** points to the ingredients in front of her by both hands  
                          and looks at **SHASHA**  
 shasha            **SHASHA** looks at **YEDA**  
 fig                    FIG 5 relates to line 031 (YEDA and SHASHA's non-verbal actions)



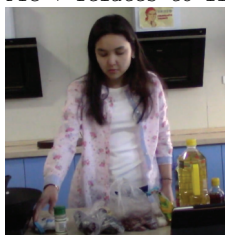
5

032                    **SHASHA**            *ni shuo shi shenme*  
                          you tell me what it is  
                          **SHASHA** looks at **YEDA**  
 shasha            **YEDA's** right hand on the salt  
 yeda                FIG 3 relates to line 032  
 fig



6

033                    (0.9)  
 yeda                **YEDA** moves closer to the first item and puts her right hand on  
                          the object  
 fig                    FIG 7 relates to line 033



7

034                    **YEDA**                *zhe ge?*  
                          this one?  
                          **YEDA** looks at **SHASHA**  
 yeda                FIG 8 relates to line 034  
 fig



035 shasha (0.2)  
 036 YEDA SHASHA nodding  
 yan  
 salt

The transition from the pre-task to the *doing language testing* sequences occurs in line 030 as SHASHA says '*just give it a try, to see how many you can remember*', while pointing to the objects on her right-hand side (FIG 3). In line 031, YEDA briefly pauses while gesturing towards the ingredients in front of her using both hands, simultaneously glancing at SHASHA as a non-verbal confirmation. In line 032 SHASHA selects YEDA as the next speaker and specifies the nature of their speech exchange system – '*You tell me what it is*' – which clearly shows that SHASHA hands over the running of their speech exchange system to YEDA, indicating that they were jointly negotiating the nature of their *doing language testing* activity and how it is organised. It is important to note here that YEDA's non-verbal action in line 031 indicates that the purpose of SHASHA's turn in line 032 is not merely a further pursuit in the absence of an SPP, but rather an expansion of her own turn in line 030 and a confirmation/nomination of how their subsequent *doing language testing* sequences will unfold.

In line 033, YEDA approaches the first item and places her right hand on the object (as depicted in FIG 7), which signifies not only her acceptance of SHASHA's invitation to engage in *doing language testing* but also serves as her self-nomination to initiate her own language testing sequences which then starts from line 036. Subsequently, YEDA asks the question '*this one?*' to confirm that they will start with '**yan**' (salt). This might be because '**yan**' (salt) is the first item of vocabulary on the system ingredient list, is on the right side and was also the first word they were tested on in the *pre-test* phase (Figure 1). After SHASHA's embodied nodding confirmation, YEDA answers correctly '**yan**' (salt) in line 036. At the start of this extract we can therefore see the nature of the speech exchange system being co-negotiated in a tentative manner, which is understandable as this is not an 'official' undertaking. In particular, YEDA in line 032 puts her hand on the salt, then in 034 looks at SHASHA and asks for confirmation as to which object to start with. SHASHA is a higher-level student than YEDA, which may explain why YEDA seems to be deferring to her at the start.

The sequential organisation from line 038 onwards is as follows. YEDA self-nominates and self-tests by picking up or pointing to the target object and then says its name, while SHASHA observes without providing any feedback. The system is not used in any way for help during this sequence. In line 038 YEDA moves to the next object ( '**hu jiao**' (black pepper)), putting

037 (1.7)  
 038 **YEDA** *emm::*  
*emm::*  
 yeda **YEDA's** right hand on black pepper  
 fig FIG 9 relates to line 038



9

039 (0.6)  
 040 **YEDA** *hū::*  
*hū::*  
 yeda **YEDA's** right hand index finger lifts up then looks down  
 fig FIG 10 relates to line 040



10



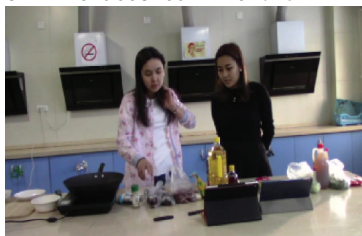
10a

041 (1.1)  
 yeda **YEDA's** right hand on eggplant  
 fig FIG 11 relates to line 041 (**YEDA's** non-verbal action)



11

042 (0.7)  
 043 **YEDA** *hū:: jiāo*  
 black pepper  
 yeda **YEDA's** right hand left on the eggplant  
 fig FIG 12 relates to line 043

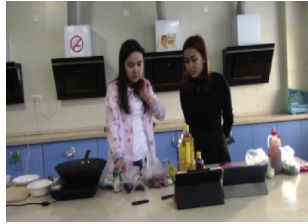


12

044

yeda  
fig

(0.5)

**YEDA's** right hand back on the eggplant  
FIG 13 relates to **YEDA's** non-verbal action

13

her hand on it. However, the short pause in the following line 037, the stretched hesitation marker '*emm*': in line 038 and another short pause in line 039 provide evidence that YEDA is potentially having problems recalling this word. Then in line 040, she utters the first word of this object in Chinese, ('*hū*'), while stretching it and holding her right-hand index finger lifted (FIGs 10 & 10a); this can be interpreted as her trying hard to remember the second word ('*jiao*') in Chinese. In line 043, she self-repairs in terms of the pronunciation and tone of the first word from '*hū*' to '*hǔ*' (though the correct pronunciation is supposed to be '*hū*') and then successfully recalls the second word '*jiao*' (pepper) with correct pronunciation.

It is important to mention here that in the CDK task the pronunciation of the correct Chinese tone of the target vocabulary item is not presented, nor practiced, nor included in the vocabulary evaluation system. This is because the focus is overwhelmingly on task completion, for which learners need to be able to identify equipment and ingredients. Since Chinese tones<sup>3</sup> are difficult for L2 learners, their inclusion would have slowed the task down considerably.

Therefore, it is quite surprising that YEDA focuses on tones and repairs herself in line 043 without any help from SHASHA or the system to produce the correct tone and pronunciation of '*hu jiao*' (black pepper) at this point. Subsequently, from line 046 to line 052, YEDA correctly produces the words for three target objects: '*qiezi*'

045

046 YEDA

(0.4)

**qiezi**  
eggplant  
(2.1)

047

048 YEDA

**er:: sheng::shengjiang**  
**er:: gin:: ginger**yeda  
fig**YEDA** picks up the ginger and then looks at the ginger  
FIG 14 relates to line 048

14

049

yeda  
fig

(1.4)

**YEDA** puts down the ginger  
FIG 15 relates to **YEDA's** non-verbal action





15

050

yeda  
fig

(0.8)

**YEDA's** right hand points at red chillies  
FIG 16 relates to **YEDA's** non-verbal action



16

051

052

**YEDA**

(1.4)

**hong::lajiao**  
**re::d chilli**

(eggplant) (line 046), '*shengjiang*' (ginger) (line 048) and '*hong la jiao*' (red chilli) (line 052) without the need for self-repair, although with some pauses.

In this case it is not merely that the learners have decided for themselves to *do language testing*, but that they have also elected to focus on pronunciation and correct an element (tones) which is not essential for the task or the post-test; the learners are adding their own learning agendas in the *doing language testing* phase.

Then in line 054, when she has already picked up the corn starch, she looks at the Chinese written name on the package. **YEDA**<sup>4</sup> attempts to say the Chinese word for 'corn' for the first time in line 054, but with an incorrect tone since it is supposed to be a falling fourth tone. She then self-repairs again in line 056 with the correct tone for '*yù*', and also

053

yeda

(1.4)

**YEDA's** right hand lifts up the corn starch

054

**YEDA**

**yùmi::**

**corn::**

**YEDA** looks at the Chinese written words on the corn starch package

055

yeda

(1.4)

056

**YEDA**

**yùmi dianfēn**

**corn starch**

yeda

looks at **SHASHA**

fig

FIG 17 relates to line 056 (**YEDA's** non-verbal action)



17

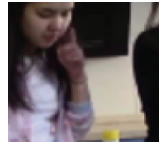


17a

057 (0.8)  
 058 YEDA **fěn**  
 flour  
 yeda YEDA's left hand mimics the 'third' tone of the word 'fěn' and puts  
 down the corn starch  
 fig FIG 18 relates to line 058 (yeda's non-verbal action)



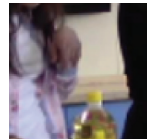
18



18a



18b



18c

059 (1.6)  
 yeda YEDA's right hand moves to the garlic  
 060 YED **dá:: suān::**  
**gar::lic::**

recalls the Chinese for 'starch'. However, she pronounces the last word 'fěn' incorrectly with a flat first tone rather than the third tone. She self-repairs again in line 058 and also embodies the trajectory of this third tone with her left hand (FIGs 18, 18a, 18b, 18c).

This sequence shows that, although these learners are not using the system for help when *doing language testing*, they are nonetheless checking the real objects in the kitchen for potential help when required. They also added their own specific L2 learning/testing focus which is not part of the CDK task, but which is a part of their broader curriculum for learning Chinese, namely learning tones. In order to complete the CDK task, it is enough for learners to be able to identify the objects from the spoken and written prompts supplied by the system, and learners do not actually need to get tones right to achieve a full 1.0 score for an item in the pre-test and post-test. We modified the previous Lexical Production Scoring Protocol-Written (Pallotti et al. 2017) (Appendix 1). It is challenging to evaluate the accuracy of foreign learners' pronunciation of Chinese tones, so we changed the rubric for a 1.0 score to 'the speaker produces the entire target lexical item without any problem in clarity or delivery' instead of the original 'the speaker produces the entire target lexical item with precision and clarity'. Nonetheless, as learners of Chinese in China, this learner realises that it is vital to use the correct tone for words to be understood and has therefore decided that it is worthwhile to spend extra time and energy on perfecting tones as part of their implementation of *doing language testing*. The learners are also creative in combining non-verbal (line 058) with verbal strategies to help themselves with learning tones. Their motivation therefore appears to be more long-term and more

oriented to the curriculum of their degree programme, rather than simply 'getting the current task done'.

In summary, during the 'self-testing' sequences in excerpt 1, the multimodal speech exchange system developed by SHASHA & YEDA involves a combination of the following components:

### 5.2. Partner-initiated testing with use of the digital system

Unlike SHASHA & YEDA, the two participants in the following excerpt, WEN & LIU (see Figure 7), involve the system as an epistemic resource for *doing language testing*. This also happens after their *pre-task* and prior to the *during-task* phase. One of the participants in effect acts more like an 'examiner' and the other more like the 'examinee', so they are *doing language testing* in a relatively more formal way. As demonstrated in Extract 2b, LIU nominates (non-verbally) the items to be tested and WEN produces their Chinese names afterwards from line 111. Lines 107 to 110 in Extract 2a clarify that before WEN's *doing language testing* sequences, LIU was the one being tested for the last two target words ('*zhi ma*' (sesame) in line 107, '*xiang cai*' (coriander) in line 109). However, when WEN does not know a word, she sometimes uses the system as an additional epistemic resource to identify and memorise the target vocabulary. This of course is not a feature of a summative test and shows that this version of *doing language testing* has learning as an aim in addition to testing; alternatively, this could be viewed as formative self-assessment. WEN & LIU are both advanced-level participants (HSK 5) (Figure 6 below) and share the same L1 but were aware that they should not speak.

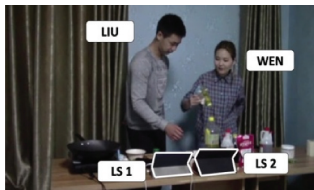
**Extract 2** '*ji bu zhu*' (I don't remember it)

Multimodal Components	Corresponding Line Numbers
a) Verbalising the target L2 words	lines 036, 043, 046, 048, 052, 054, 056, 058, 060
b) Deictic reference to items in the kitchen by eye gaze, pointing at, touching, or moving them	lines 030, 032, 034, 038, 040, 041, 043, 044, 048, 049, 050, 053, 054, 056, 058, 059
c) Addressing each other using verbal and/or non-verbal means	lines 030, 032, 034, 056
d) Physical embodied actions	lines 030, 032, 034, 038, 040, 041, 043, 044, 046, 048, 049, 050, 053, 054, 056, 058, 059

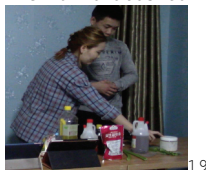
Figure 6. SHASHA & YEDA multimodal components.

	WEN	LIU
<b>HSK Level</b>	HSK 5 (advanced)	HSK 5 (advanced)
<b>Ethnicity</b>	Kyrgyzstani	Kyrgyzstani
<b>Gender</b>	Female	Male
<b>Age Group</b>	18-25	18-25
<b>University Programme</b>	Law (MA)	Chinese Language and Literature (MA)
<b>Arrived in China</b>	2016	2015
<b>Length of Stay in China</b>	3 years	4 years

Figure 7. WEN&LIU's information.



- 107      **LIU**      *zhima*  
                         sesame  
                         **LIU** looks at the sesame  
liu              **WEN** points at the sesame with her right hand index finger and looks at it  
wen              FIG 19 relates to line 107  
fig



- 108      (0.5)  
109      **LIU**      *xiang cai*  
                         coriander  
                         **LIU** looks at the coriander  
liu              **WEN** points at the coriander with her right hand index finger  
wen              FIG 20 relates to line 109  
fig



110 WEN *en, ke yi.*  
em-hmm, ok.



The above Extract 2a shows LIU's *doing language testing* sequences while he was being tested for the last two words, and his co-participant WEN was pointing at the target objects. A confirmation token 'ok' in line 110 marks the end of LIU's testing sequences.

From line 111 in the Extract 2b below, the two learners switch over using both verbal and non-verbal means.

111 LIU *lai, ni lai*  
come here, it's your turn  
wen WEN looks at the ingredients on the table  
liu LIU's right arm out on WEN's left arm and  
walks to the right side of him  
fig FIG 22 relates to line 111



112 (1.2)  
113 WEN *.hhh*  
*.hhh*  
wen WEN walks to the left  
fig FIG 23 relates to line 113



114 (1.0)  
liu LIU at the right side of WEN, looks down  
wen WEN at the left side of LIU, looks down and smiles  
fig FIG 24 relates to line 114



24

115

liu  
wen  
fig

(0.6)

LIU's right hand points at the red chilli

WEN looks at the red chilli

FIG 25 relates to line 115



25

116

117

WEN

(0.6)

la jiao

chilli

118

(1.2)

liu  
wen  
fig

LIU's right hand moves onto the garlic

WEN looks at the garlic

FIG 26 relates to line 118



26

119

WEN

suan

garlic

As shown in the above Extract 2b, it is noticeable that LIU first makes an announcement in line 111 – '*lai ni lai*' (come here, it's your turn), that nominates WEN as the next speaker to be tested. Then they switch positions as LIU wants to point to specific objects for which WEN should produce their Chinese names. WEN responds with laughter and moves to the left of LIU (see from line 113 to line 114, FIGs. 16 & 17). Unlike SHASHA & YEDA in *Excerpt 1*, who were trying to select items in their language testing phase following the order of the 'pre-test' phase, LIU randomly points at the red chilli as the first test item (FIG 18 in line 115). WEN answers immediately, displaying her ability to produce '*la jiao*' (chilli) in the following line 117. The same sequence is repeated from line 118 to line 119 for '*da suan*' (garlic), even though WEN actually offers an alternative answer: '*suan*'. This is however acceptable because '*suan*' is normally used for garlic in Chinese, especially in daily life situations.

120

liu (2.3)  
 fig LIU's right hand index finger points at the corn starch  
 FIG 27 relates to LIU's non-verbal action



27

121

WEN °ji bu zhu °  
 °I don't remember it°  
 wen WEN's right hand picks up the corn starch bag and looks at the  
 fig Chinese written name on the corn starch bag  
 FIG 28 relates to line 121 (wen's non-verbal action)



28

122

wen (1.7)  
 fig WEN holds the corn starch with both hands and takes a close  
 look at the Chinese written name on it  
 FIG 29 relates to wen's non-verbal action



29

123

WEN yu : :  
 /yu : :  
 (0.8)

124

wen WEN puts down the corn starch

125

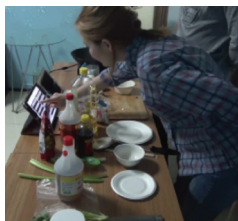
WEN ei ;  
 ei ↑  
 wen WEN looks at the system  
 fig FIG 30 relates to line 125



30

126

(0.9)  
 wen WEN's right hand index finger points at the system and searching  
 fig FIG 31 relates to WEN's non-verbal action

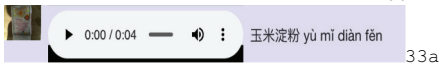


31

127

(0.2)

wen **WEN**'s right hand index finger points at 'corn starch' on the list of the **system**  
 fig FIG 32 relates to **WEN**'s non-verbal action  
 FIG 33 relates to the digital system's interface where **WEN** points at



128

(0.8)

129

**WEN** *yumi:: yumi:: dianfen*  
 corn:: corn:: starch  
 wen **WEN** stands still and looks at the **system**  
 fig FIG 34 relates to line 129



130

(1.5)

131

**WEN** *yumi dianfen*  
 corn starch  
 wen **WEN**'s right hand index finger points at the corn starch  
 fig FIG 35 relates to line 131 (wen's non-verbal action)



131

**WEN** *yumi dianfen*  
 corn starch  
 wen **WEN**'s right hand index finger points at the corn starch  
 fig FIG 35 relates to line 131 (wen's non-verbal action)



132

(0.3)

liu **LIU** nodding



After that in line 120 LIU points at the next object – ‘*yumi dianfen*’ (corn starch) – while WEN displays her lack of knowledge explicitly in the following line, by saying ‘*ji bu zhu*’ (I don’t remember it). At the same time, she also picks up the corn starch bag and studies the Chinese written name on it for a while (see FIGs 28 & 29). Then in line 123, she attempts to read and pronounce the first word ‘*yu*’ but then puts it down and switches her focus to the system, searching through the ingredient list. In line 125, WEN says ‘*ei↑*’ which is a change-of-state token in Chinese representing a realisation that something is not right. This indicates that WEN has a problem reading out the Chinese characters for at least the last three words ‘*mi dian fen*’. She searches for a while and finally in line 129, reads out the full name of this target object (FIG 34) despite some stretching, repetition and hesitation at the beginning. She then repeats it without stretching the sound in line 131, which can be understood as her self-confirming that she has now achieved a full understanding of this word. This is other-confirmed by LIU’s embodied action in line 132, nodding as a way of expressing positive evaluation in the ‘examiner’ role. It is also noticeable that during WEN’s epistemic search sequences from line 122 to line 131, LIU initiates no repair or feedback of any kind.

In the above sequence we can see a number of differences between the speech exchange system developed here and the pre-test with the researcher. The testee is able to refer to both a real object (corn starch) and to the system to obtain help. Also, LIU nods to confirm correctness, which the researcher does not do in the pre-test.

133

liu  
fig

(0.3)

**LIU**’s right hand on chopping board  
FIG 36 relates to **LIU**’s non-verbal action

36

134

**WEN**

(0.3)

**an ban**  
chopping board

136

liu  
fig

(0.8)

**LIU**’s right hand index finger on the lid of olive oil  
FIG 37 relates to **LIU**’s non-verbal action

37

137→

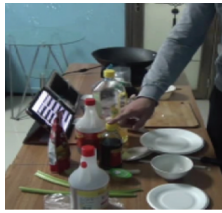
**WEN****guo!**  
pass!

138

liu  
fig

(2.3)

**LIU**’s right hand index finger moves onto the sesame oil  
FIG 38 relates to **LIU**’s non-verbal action



38

139 WEN *zhima:: zhima you*  
sesame:: sesame oil  
(2.4)

140 liu LIU's right hand index finger on the chicken stock  
fig FIG 39 relates to LIU'S non-verbal action



39

141 WEN *mei you zhe ge ba?*  
it doesn't have this one, does it?

wen WEN's left hand on the chicken stock and looks at the **system**  
fig FIG 40 relates to line 141 (wen's non-verbal action)



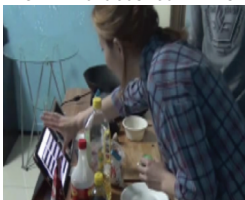
40

Then they move to the next object '*an ban*' (chopping board), and like the first two vocabulary items, WEN produces this accurately and rapidly. However, when in line 136 LIU moves to the next object ('*gan lan you*' or olive oil) WEN surprisingly answers '*guo*' (pass) without making any attempt to either ask LIU for help, nor to search on the system like she did for corn starch (line 127). LIU does not offer to help WEN learn this vocabulary item (just like the researcher in the pre-test) and moves to the next prompt in line 138, which WEN then answers correctly with slight hesitation '*zhima: zhima you*' ('sesame: sesame oil'). When the testing moves onto the next object, '*gu ti ji tang*' (chicken stock), WEN neither searches the system nor says '*guo*' (pass), but queries to LIU whether this item is on the system's vocabulary learning/testing list – '*mei you zhe ge ba?*' (it doesn't have this one, does it?) in line 141. This is not WEN challenging LIU or the Linguacuisine system, but more like a self-questioning/self-repair in terms of her own forgetting, functioned as an information request to LIU. She then checks this herself by leaning forward to the system, scrolling up and down to search for this object.

It is important to note here that in the current research setting without professionals present, learners' interactional patterns can change relatively, when it has not to be a certain type of interaction nor a copy of normal language testing with professionals present. What happened here is something unusual and not the same subjective constraints with formal tests.

142

(2.5)  
 wen **WEN** leans forward, right hand scrolls up and down on the **LS 2**  
 liu **LIU** stands still and looks at the **system**  
 fig **FIG 41** relates to **WEN's** non-verbal action



41

143

**WEN** *zai zhe li ma?*  
 is it here?

144

**WEN** continues surveying the **system**  
**LIU** searches together with **WEN** on the **system**  
 (0.3)

145

liu **LIU's** right hand index finger points at the 'chicken stock' on the list of **system**  
 fig **FIG 42** relates to **LIU's** non-verbal action  
**FIG 43** relates to the system interface where **LIU** points at



42



43a

146

**WEN** °a:ː° (0.9) °guti ji tang°  
 °ah:ː° (0.9) °chicken stock°

wen **WEN** picks up the chicken stock and looks at it  
 fig **FIG 44** related to line 146 (**WEN's** non-verbal action)



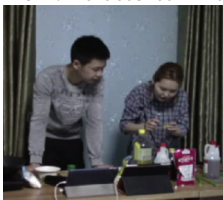
44

147

148

**LIU** (0.9) *ji tang*  
 chicken broth

liu **LIU** look left at the chicken stock  
 fig **FIG 45** relates to line 148 (**WEN's** non-verbal action)



45

149

150

**WEN** (1.9) *ji tang*  
 chicken broth

151

152

**WEN** en  
 yes

wen **WEN** puts down the chicken stock

As LIU does not respond, WEN reframes her question as requesting confirmation of whether it is on the list in line 143 – ‘*zai zhe li ma*’ (is it here?). This, however, still elicits no response from LIU. WEN searches on the system for a quite long time until LIU finally decides to help by pointing at the target vocabulary on the system list (line 145, FIG 35). This action successfully upgrades WEN’s epistemic progression by not only answering her former question (is it on the list?) but also locating it on the system as a resource for learning. This helps WEN to find the name of the object. In line 146, she starts with a ‘*ah*’ and then pronounces the word correctly, finally displaying her full grasp of this item. LIU’s action therefore actively helped WEN succeed in changing her epistemic state, a move which the researcher does not make in the pre-test. LIU also repeats the last two words in line 148, which is again a move not undertaken by the researcher in the pre-test. We can note that LIU steps out of the ‘examiner’ role twice in this extract. In both cases WEN was clearly experiencing extended trouble in locating the word for the item, thus slowing the overall progress of the session down.

In summary, the multimodal speech exchange system between LIU & WEN in Excerpt 2 involves a combination of the following components:

Whilst components a)-d) are the same as in Extract 1 (Figure 8), the use of the system has added components e), f) and g) in Extract 2.

We compared how the learners sequenced the items and found that the learners in Extract 1 followed exactly the same pre-test sequence of names as in the pre-test (Nos. 1 to 7 in Figure 9 below). By contrast, WEN & LIU in *Extract 2* pointed at random objects on the table in front of them (Nos. 5, 6, 12). The pre-test sequence of target object names are shown below in a photo of the researcher’s notebook:

Multimodal Components	Corresponding Line Numbers
a) Verbalising the target L2 words	lines 117, 119, 123, 129, 131, 135, 139, 146, 148, 150
b) Deictic reference to items in their kitchen by eye gaze, pointing at, touching, or moving them	lines 111, 115, 118, 120, 121, 122, 124, 125, 126, 131, 133, 136, 138, 140, 141, 146, 148, 152
c) Addressing each other using verbal and/or non-verbal means	lines 111, 132, 137, 141, 143, 148, 152
d) Physical embodies actions	lines 111, 113, 114, 115, 118, 120, 121, 122, 124, 125, 126, 127, 129, 131, 132, 133, 136, 138, 140, 141, 142, 143, 144, 145, 146, 148, 152
e) Matching the Chinese name of the target item on the system screen with the object in their kitchen	lines 121, 122, 123, 131, 146, 148, 150
f) Reading the Chinese name of the target item on the system’s screen	lines 127, 129, 146
g) Pressing buttons on the system	lines 126, 127, 142, 144

Figure 8. LIU & WEN multimodal components.

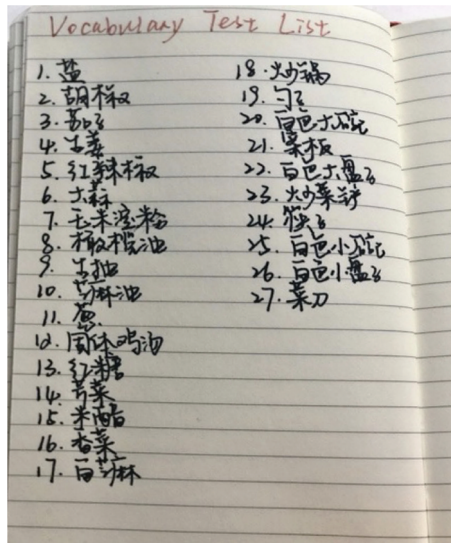


Figure 9. Pre-test sequence of test items.

### 6. Learner motivations for doing language testing: interviews

Having established the basic nature and organisation of the *doing language testing* side sequences, we now ask why exactly the learners engage in these, since they were not obligatory, nor indeed suggested by the researcher or task/test structure in any way. Indeed, the sessions were time-consuming for the learners as it was, without them adding additional sequences. In order to answer this question, we examined post-hoc interviews (in Chinese) with all of the 12 learners who had *done language testing*. The 4 learners featured in the extracts said that a major motivation for *doing language testing* was to have an improved learning outcome during the main task (Figure 10):

**Extract 3**

YEDA:	well, em:: I th::nk (0.3) that can help us be more familiar with all of the things
SHASHA:	in that way, I believe our cooking process will go very smoothly
WEN:	I also thi::nk doing this could help us (0.2) help us:: remember these words in a better way, (0.4) like some of those words (0.3) I've never seen them, so I (.) I:: could pay more (.) mo::re attention to them during the cooking session.
LIU:	I thin::k (0.3) doing that testing befo::re our cooking session, the learning outcome (0.2) can (.) ca::n be better!

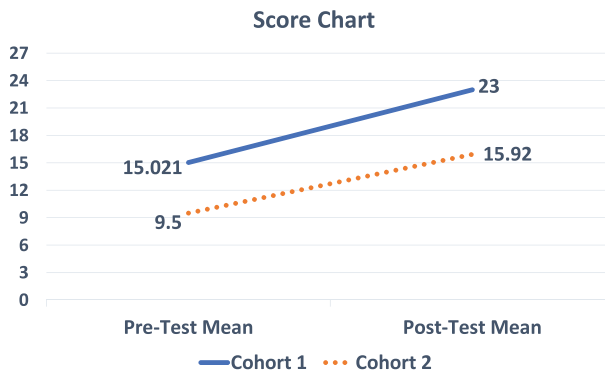
Figure 10. Participants' interview data.

None of the 12 students mentioned the prospect of getting a better post-test result as a motivation, nor did anyone say that the pre-test motivated them to *do language testing*. However, WEN mentioned that their experience in the Chinese L2 classroom had been that the teacher would check on vocabulary learning before the main activity, suggesting that replicating previous classroom experience (rather than testing experience) was a motivation for some learners. We also considered how much extra time students were investing as this may reflect their degree of motivation. On average the 12 learners added 3.26 minutes to their sessions (average length 43.06 minutes) by *doing language testing*.

## 7. Evidence of vocabulary learning: pre-test and post-test

We also asked what the possible association might be between *doing language testing* and vocabulary gain. Might this voluntary extra work bear extra fruit in terms of enabling enhanced vocabulary acquisition? To investigate this, we divided the test results of the whole cohort of learners into two sub-cohorts. Sub-cohort 1 consisted of the 12 students *doing language testing* and cohort 2 of the 60 students who did not do so. We compared pre-test and post-test means for both sub-cohorts, which then produced a mean for improvement of vocabulary score – see Figure 11 below. It is worth noting that our emphasis here is not on how students' language proficiency level may affect the initiation of *doing language testing* sequences. Rather, our focus centres on the improvements arising from their *doing language testing* sequences.

This figure revealed a statistically significant difference, with cohort 1 demonstrating higher scores in both tests. To account for the initial variance in pre-test means, we recalculated scores for proportional gains, illustrating in Figure 12 that cohort 1 experienced a growth of 29.56%, while cohort 2 had an average gain of 23.78%. Statistical results are as follows:



**Figure 11.** Pre-test and post-test mean of the two cohorts.

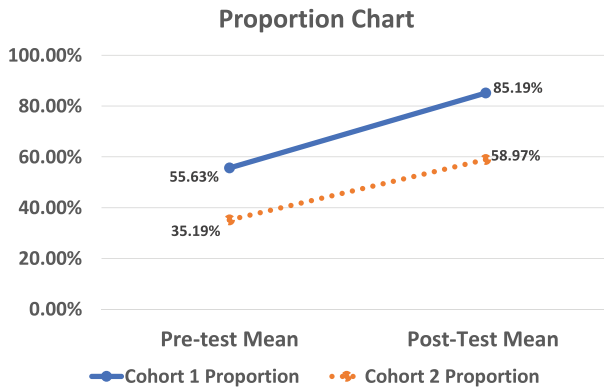


Figure 12. Proportion chart of the two cohorts.

	<b>Cohort 1</b>	<b>Cohort 2</b>
<b>N</b>	12 (6 pairs)	60 (30 pairs)
<b>Pre-test mean</b>	15.02	9.5
<b>Post-test mean</b>	23	15.92
<b>P value</b>	p-value = 0.00 <b>&lt;0.05</b>	p-value = 0.00 <b>&lt;0.05</b>
<b>T value</b>	10.90	10.88
<b>Effect Size (Cohen’s D)</b>	2.30	0.93

Figure 13. Pre-test and post-test means of the two cohorts.

An initial 5.5 difference in pre-test scores (Figure 13) indicated that the *doing language testing* sub-cohort had a significantly greater initial familiarity with the vocabulary items. Did that sub-cohort start at a higher level of overall proficiency? Confirming this, HSK proficiency scores showed a higher mean overall proficiency level for the language testing sub-cohort (7.979 vs. 6.42).

Additionally, a t-test comparing improvement confirmed that the *doing language testing* sub-cohort had significantly greater improvement ( $t = 1.703$ , Figure 14).

In Figure 14, we examine the disparity in improvement between cohort 1 and cohort 2. Cohort 1 displays a more substantial enhancement of 7.98, while cohort 2 shows a less pronounced enhancement of 6.36. The enhancement observed in cohort 1 exhibits a statistically significant increase compared to cohort 2 ( $p\text{-value} < 0.05$ ). Our analysis provides statistical confirmation of the distinctions between these two cohorts.

### 8. Conclusions

This study provides evidence that some learners perceived the practice of *doing language testing* to be of practical help to them, to the extent that they were prepared to invest, of

	<b>im_cohort 1</b>	<b>im_cohort 2</b>
<b>Mean</b>	7.98	6.36
<b>Variance</b>	6.44	20.53
<b>Observations</b>	12	60
<b>df</b>	27	
<b>t Start</b>	1.72	
<b>P (T&lt;=t) one tail</b>	0.048	
<b>t critical one tail</b>	1.70	
<b>Effective Size (Cohen's D)</b>	0.44	

**Figure 14.** T-Test results of the improvements of two cohorts.

their own volition, unscheduled time and effort into it. The overall picture we have obtained from interview and test data is that a) *doing language testing* contributed to their enhanced gain in vocabulary learning and relatively higher post-test scores b) the learners' major motivation for *doing language testing* was to improve their own learning and task performance, rather than a love of testing per se. Learners *doing language testing* believed that the extra time and effort would be worth it in terms of enhanced learning experience and results; the score differential suggested that this was indeed the case. The literature reviewed in section 2 is of peer assessment as organised by teachers, whereas this is the first study of learners initiating and organising assessment themselves for their own benefit.

In terms of learning and teaching implications, the study clearly shows that some students are willing and motivated to spend extra time and effort on peer/self-testing of vocabulary items prior to a main task as they believe this will enhance their task performance. All 12 learners selected the same slot (between pre-task and main task) as the perfect time for *doing language testing*, which has clear implications for effective TBLT practice. These learners were able to organise the interaction for themselves in the way which best suited them and were able to introduce their own interests which were external to the allocated task, for example a focus on tones.

In terms of methodological implications, Seedhouse (2022) suggested that a mixed methods approach including CA is necessary to portray L2 learning fully. Our study demonstrates and supports that a CA approach was effective in discovering the totally unexpected phenomenon of *doing language testing* in our context and in then describing and analysing its sequential organisation. This is because CA as a methodology is designed to reveal the complexity and fluidity of spoken interaction, and has been extremely successful and popular as a methodology for the analysis of the complex organisation of ordinary conversation (Sacks, Schegloff, and Jefferson 1974), of varieties of institutional interaction (Drew and Heritage 1992) and of spoken interaction in a huge



range of settings (Seedhouse 2022). However, an idiosyncratic problem with language learning talk is that it adds an extra layer of complexity to the analysis, as language itself becomes both the object and vehicle of language learning talk (Seedhouse 2004). Therefore, interview and test data were then necessary to explore the motivations of learners in giving themselves extra work, as well as the relationship between the phenomenon and learning gains. The *triangulation* mixed methods approach was used to provide 3 different but complementary perspectives on the phenomenon in order to illuminate its various characteristics. We have obtained a more rounded perspective on *doing language testing*, throwing light on how learners do it, why they elect to do it and what the consequences of doing it may be for their learning. A limitation is that only 12/72 learners chose to do this. Furthermore, these learners were at relatively high proficiency levels, so we may have found that higher-level, more motivated learners decided to spend extra time *doing language testing*, successfully increasing their proficiency. This may turn out to be a relatively niche practice, but on the other hand the evidence suggests that *doing language testing* is likely to prove to be an effective strategy for L2 learners.

We have described the practice of *doing language testing* in one specific setting only, solely in relation to pairs of learners using an app without a teacher present. We noted variation in terms of the sequential organisation and strategies used for *doing language testing* by the two different pairs. Learners did not simply follow the sequential organisation which they had experienced in the pre-test with the researcher. It may be that *doing language testing* occurs in many other formal and informal settings for language learning, where the sequential organisation may be very different; this is clearly an area for future research. What can the ways that L2 learners decide to test themselves (self-testing) or each other (peer-testing) tell us about the best ways to do L2 learning, teaching and testing? This study has demonstrated ways in which we can gain insights into learners' own perspectives on effective learning and on the strategies they actually adopt themselves in practice to improve their own learning experience. This area may be a 'coral garden' ripe for future research. We certainly never expected to uncover this phenomenon in our study.

## Notes

1. This article is based on the completed PhD project by Ren, S. (2022). *How Paired Learners of L2 Chinese Manage Intersubjectivity and Epistemic Status to Complete Technology-Mediated Tasks: A Multimodal Conversation Analytic Study*. Doctoral Thesis. Newcastle University. The Newcastle University Ethics Committee for Humanities and Social Sciences approved this project on 3rd Jan, 2018. All of the subjects have provided informed consent in writing for publication.
2. Readers can try the task themselves on <https://enacteuropa.com/?q=node/281>.
3. The Chinese language has four tones (the first tone is a flat tone, second is a rising tone, third is a dip tone and fourth is a falling tone), but there is no teaching of tones in the CDK task, so this focus on tone learning by the learners is entirely self-nominated and incidental. It is in fact very difficult for L2 learners to pronounce each tone correctly every time, which is the reason why tones are not a part of the task or assessment. This study did not attempt to focus on tones for the following reasons: 1) In Chinese/Mandarin, words are either single syllable or are made by putting together two single-syllable words. Each syllable has its own meaning and can have one tone that defines the meaning of the syllable or the word. Tones are used to differentiate words from each other, like consonant and vowel combinations in English. Correct pronunciation regarding tones is therefore important, as is the ability to distinguish

- tones while listening; 2) However, it can be very challenging for learners whose L1 is non-tonal to not only listen to and distinguish the Chinese tones but also product them precisely; 3) it is also very difficult for us as analysts or researchers to evaluate whether they have achieved the 'native-like' or precise standard of the tones. Therefore, tones remain unmarked in our transcripts; they are only marked when participants decided to learn/repair them.
4. On the corn starch package, there is no tone marked, only Chinese words written in pinyin, namely '*yu mi dian fen*' (corn starch flour).

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

The Linguacuisine project was supported by the European Union Erasmus Plus Programme 2016-18 under a grant of €324K number [2016-1-UK01-KA204-024254]. Partners: Newcastle University, Hellenic Open University, Workers Educational Association, Action Foundation, University of Modena and Reggio-Emilia. The ENACT project was co-funded by the European Commission, Erasmus Key Action 2 Strategic Partnerships for Higher Education 2019-22, project number: [2019-1-UK01-KA203-061567].

## Data availability statement

Video data for the two excerpts are available:

Excerpt 1 (SHASHA & YEDA): <https://youtu.be/eb0oZCdZKGk>.

Excerpt 2 (WEN & LIU): <https://youtu.be/eBWAJu5w8K8>.

## ORCID

Simin Ren  <http://orcid.org/0000-0002-7678-1427>

Paul Seedhouse  <http://orcid.org/0000-0002-0099-3167>

## References

- Andrade, M. S. 2016. "Effective Organizational Structures and Processes: Addressing Issues of Change." *New Directions for Higher Education* 2016 (173): 31–42. <https://doi.org/10.1002/he.20177>.
- Bachman, L. F. 1989. "The Development and Use of Criterion-Referenced Tests of Language Ability in Language Program Evaluation." *The Second Language Curriculum* 242–258.
- Bachman, L. F., and A. S. Palmer. 1989. "The Construct Validation of Self-Ratings of Communicative Language Ability." *Language Testing* 6 (1): 14–29.
- Brown, H. D., and P. Abeywickrama. 2010. *Language Assessment: Principles and Classroom Practices*. Vol. 10. White Plains, NY: Pearson Education.
- Bryan, C., and K. Clegg. 2019. Conclusion: Resilience, resourcefulness and reflections. In *Innovative Assessment in Higher Education*, edited by C. Bryan and K. Clegg, 240–249. New York: Routledge. <https://doi.org/10.4324/9780429506857>.
- Chang, C., and H. C. K. Lin. 2020. "Effects of a Mobile-Based Peer-Assessment Approach on Enhancing Language-Learners' Oral Proficiency." *Innovations in Education and Teaching International* 57 (6): 668–679. <https://doi.org/10.1080/14703297.2019.1612264>.
- Cheng, W., and M. Warren. 2005. "Peer Assessment of Language Proficiency." *Language Testing* 22 (1): 93–121. <https://doi.org/10.1191/0265532205lt298oa>.

- Creswell, J. W., and V. L. Plano Clark. 2017. *Designing and Conducting Mixed Methods Research*. Sage Publications.
- Drew, P., and J. Heritage. 1992. "Contested Evidence in Courtroom Cross-Examination: The Case of a Trial for Rape." *Talk at Work: Interaction in Institutional Settings*.
- Ellis, R. 2001. "Introduction: Investigating Form-Focused Instruction." *Language Learning* 51 (s1): 1–46. <https://doi.org/10.1111/j.1467-1770.2001.tb00013.x>.
- Ellis, R. 2003. *Task-Based Language Learning and Teaching*. Oxford: Oxford University Press.
- Falchikov, N. 1995. "Peer Feedback Marking: Developing Peer Assessment." *Innovations in Education and Training International* 32 (2): 175–187. <https://doi.org/10.1080/1355800950320212>.
- Freeman, M. 1995. "Peer Assessment by Groups of Group Work." *Assessment & Evaluation in Higher Education* 20 (3): 289–300.
- Fulcher, G., and F. Davidson. 2007. *Language Testing and Assessment*. London and New York: Routledge.
- Gardner, R., and J. Wagner, eds. 2004. *Second Language Conversations*. Continuum, London/New York: A&C Black.
- González-Lloret, M., and L. Ortega. 2014. "Towards Technology-Mediated TBLT." *Technology-Mediated TBLT: Researching Technology and Tasks* 6: 1–22.
- Hoffman, B. 2019. "The Influence of Peer Assessment Training on Assessment Knowledge and Reflective Writing Skill." *Journal of Applied Research in Higher Education* 11 (4): 863–875. <https://doi.org/10.1108/JARHE-01-2019-0004>.
- Jefferson, G. 1972. "Side Sequences." In *Studies in Social Interaction*, edited by D. Sudnow, 295–338. New York: Free Press.
- Kasper, G., and S. J. Ross. 2007. "Multiple-Questions in the Oral Proficiency Interview." *Journal of Pragmatics* 39 (11): 2045–2070. <https://doi.org/10.1016/j.pragma.2007.07.011>.
- Kormos, J. 1999. "Simulating Conversations in Oral-Proficiency Assessment: A Conversation Analysis of Role Plays and Non-Scripted Interviews in Language Exams." *Language Testing* 16 (2): 163–188.
- Kurhila, S., and L. Kotilainen. 2020. "Student-Initiated Language Learning Sequences in a Real-World Digital Environment." *Linguistics & Education* 56 (April): 100807. <https://doi.org/10.1016/j.linged.2020.100807>.
- Long, M. 2015. *Second Language Acquisition and Task-Based Language Teaching*. Oxford: Wiley-Blackwell.
- Meletiadou, E. 2021. "Exploring the Impact of Peer Assessment on EFL Students' Writing Performance." *IAFOR Journal of Education: Language Learning in Education* 9 (3): 77–95. <https://doi.org/10.22492/ije.9.3.05>.
- Meletiadou, E., and D. Tsagari. 2016. "The Washback Effect of Peer Assessment on Adolescent EFL Learners in Cyprus." In *Classroom-Based Assessment in L2 Contexts*, edited by D. Tsagari, 138–160. Newcastle Upon Tyne: Cambridge Scholars Press.
- Olivier, P. G. X., A. Monk, and J. Hoey. 2009. "Ambient Kitchen: Designing Situated Services Using a High Fidelity Prototyping Environment." In *Proceedings of the 2nd international conference on pervasive technologies related to assistive environments*, 1–7.
- Orsmond, P., and S. Merry. 1996. "The Importance of Marking Criteria in the Use of Peer Assessment." *Assessment & Evaluation in Higher Education* 21 (3): 239–250. <https://doi.org/10.1080/0260293960210304>.
- Pallotti, G., N. Niemants, and P. Seedhouse. 2017. "Vocabulary learning in a real-world digital environment." *Task-based language learning in a real-world digital environment: The European digital kitchen* 207–230.
- Patri, M. 2002. "The Influence of Peer Feedback on Self- and Peer-Assessment of Oral Skills." *Language Testing* 19 (2): 109–131. <https://doi.org/10.1191/0265532202lt224oa>.
- Peng, J. 2010. "Peer Assessment in an EFL Context: Attitudes and Correlations." In *Selected Proceedings of the 2008 Second Language Research Forum*, edited by M. T. Prior, et al., 89–107.
- Pope, N. 2001. "An Examination of the Use of Peer Rating for Formative Assessment in the Context of the Theory of Consumption Values." *Assessment & Evaluation in Higher Education* 26 (3): 235–246.

- Preston, A., M. Balaam, P. Seedhouse, S. Kurhila, L. Kotilainen, A. Rafiev, D. Jackson, and P. Olivier. 2015. "Can a Kitchen Teach Languages? Linking Theory and Practice in the Design of Context-Aware Language Learning Environments." *Smart Learning Environments* 2 (1): 1–19.
- Reinholz, D. 2016. "The Assessment Cycle: A Model for Learning Through Peer Assessment." *Assessment & Evaluation in Higher Education* 41 (2): 301–315. <https://doi.org/10.1080/02602938.2015.1008982>.
- Sacks, H., E. A. Schegloff, and G. Jefferson. 1974. "A Simplest Systematics for the Organization of Turn Taking for Conversation. In *Studies in the Organization of Conversational Interaction*, edited by J. Schenkein. 7–55. Academic Press.
- Sawchuk, P. 2003. *Adult Learning and Technology in Working-Class Life*. Cambridge: Cambridge University Press.
- Seedhouse, P. 2004. *The Interactional Architecture of the Language Classroom: A Conversation Analysis Perspective*. Malden, MA: Blackwell.
- Seedhouse, P. ed. 2017. *Task-Based Language Learning in a Real-World Digital Environment: The European Digital Kitchen*. Vol. 4. Bloomsbury Publishing.
- Seedhouse, P. 2022. "Transitioning from Conversation Analysis to Mixed Methods." *Language Teaching* 1–12. <https://doi.org/10.1017/S0261444822000027>.
- Seedhouse, P., P. Heslop, A. Kharrufa, S. Ren, and T. Nguyen. 2019. "The Linguacuisine Project: A Cooking-Based Language Learning Application." *The EuroCALL Review* 27 (2): 75–97. <https://doi.org/10.4995/eurocall.2019.13663>.
- Seedhouse, P., and F. Nakatsuhara. 2018. *The Discourse of the IELTS Speaking Test: Interactional Design and Practice*. Cambridge: Cambridge University Press.
- Seedhouse, P., A. Preston, P. Olivier, D. Jackson, P. Heslop, M. Balaam, A. Rafiev, and M. Kipling. 2014. "The European Digital Kitchen Project." *Bellaterra Journal of Teaching and Learning Language and Literature* 7 (1): 0001–16.
- Seedhouse, P., A. Preston, P. Olivier, D. Jackson, P. Heslop, T. Plötz, M. Balaam, and S. Ali. 2013. "The French Digital Kitchen: Implementing Task-Based Language Teaching Beyond the Classroom." *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)* 3 (1): 50–72.
- Thomas, M., and H. Reinders. 2010. "Deconstructing tasks and technology." *Task-Based Language Learning and Teaching with Technology*, edited by M. Thomas and H. Reinders, 1–16. London: Continuum International Publishing Group.
- Topping, K. J. 2009. "Peer Assessment, Theory into Practice." *Theory into Practice* 48 (1): 20–27. <https://doi.org/10.1080/00405840802577569>.
- Topping, K. J. 2017. "Peer Assessment: Learning by Judging and Discussing the Work of Other Learners." *Interdisciplinary Education and Psychology* 1 (1). <https://doi.org/10.31532/InterdiscipEducPsychol.1.1.007>.
- Topping, K. J. 2018. *Using Peer Assessment to Inspire Reflection and Learning*. Routledge. <https://doi.org/10.4324/9781351256889>.
- Weiser, M. 1999. "The Computer for the 21st Century." *ACM SIGMOBILE Mobile Computing and Communications Review* 3 (3): 3–11.
- Youn, S. J., and Burch, R. 2020. "Where Conversation Analysis Meets Language Assessment: Toward Expanding Epistemologies and Validity Evidence." *Papers in Language Testing and Assessment* 9. <https://doi.org/10.58379/DYKI2461>.
- Zamora, Á., J. M. Suárez, and D. Ardura. 2018. "Error Detection and Self-Assessment as Mechanisms to Promote Self-Regulation of Learning Among Secondary Education Students." *Journal of Educational Research* 111 (2): 175–185. <https://doi.org/10.1080/00220671.2016.1225657>.

## Appendices

### 1 Appendix

#### Modified Lexical Production Scoring Protocol-Adapted for CDK

Score	Speaker Spoken Production
0.00 points	The speaker <b>says nothing</b> at all or states that s/he is unable to answer.
0.25 points	The speaker <b>makes an attempt to name the target object</b> which is unintelligible and is very difficult to understand in relation to the target object.
0.50 points	The speaker <b>produces the target lexical item partially</b> , or in a way which can only be understood to relate to the target object with some difficulty, with a major problem in pronunciation and/or clarity. Or the speaker tried to describe the object rather than name it.
0.75 points	The speaker <b>produces the entire target lexical item</b> in an intelligible way, but with a <b>minor problem</b> in pronunciation and/or clarity, or in delivery.
1.00 points	The speaker produces <b>the entire target lexical item</b> without any problem in clarity or delivery. ( <b>note:</b> the original criterion in Pallotti et al. (2017) – ‘the speaker produces the entire target lexical item with precision and clarity’).

### 2 Appendix

#### Transcription Conventions

(0.2)	The tenths of a second between utterances. Interval between utterances
(.)	A micro-pause (1 tenth of a second or less)
:	Sound extension of a word (more colons demonstrate longer stretches)
∞	Talk that is quieter than surrounding talk
↑	Rising intonation
hhh	Audible aspiration
.hh	Audible inhalations
fig.	number of the screenshot (from video recordings)