Ludic Language Pedagogy

https://www.llpjournal.org

Pedagogical considerations for teaching with games: Improving oral proficiency with self-transcription, task repetition, and online video analysis

James York*

Tokyo Denki University, College of Science and Engineering

ARTICLE INFO

Article history:
Received: March 11, 2020
Revised: September 05, 2020
Accepted: October 07, 2020
Published: October 20, 2020

Keywords:
Board games
CAF model
Smartphones
TBLT
Transcription
YouTube

Peer reviewers:
Vita Kogan
Frederick Poole

KEY POINTS

Background: Board games were utilized with rigorous pre and post-play activities in an EFL classroom context to promote L2 development.

Aim: The aim of this paper is to assess whether the post-play activities improved participants' speaking skill.

Methods: Using complexity, accuracy and fluency measures, as well as reference to learners' completed workbooks, output performance was analysed quantitatively.

Results: Output accuracy was significantly improved in the second gameplay session, however complexity and fluency dimensions were not affected. A number of common morphosyntactic errors were also left unnoticed.

Conclusion: Whilst student-driven post-tasks aided learners' L2 development; particularly along the fluency dimension, additional teacher instruction is required.

Tweet synopsis

Learners improved their speaking skills over 2 gameplay sessions with #transcription and error-analysis tasks. However, many errors were left unnoticed. Supporting gameplay with student and teacher-driven tasks can aid learning in ludic language pedagogy!

View at the LLP website: https://llpjournal.org/2020/10/20/j-york-transcription.html

* Corresponding author. *Email address: jamesyorkjp@gmail.com (Dr. James York)*

Introduction

This paper is a continuation of the work I have been doing in my classroom context. The context--a private university in Japan--dictates that I am to teach communicative language skills from a weak CLT perspective as is typical in many universities in Japan (see Johnson, Lyddon, Selman, & Nelson, 2015). By this I mean that the class is designed around language functions and the systematic teaching of forms via a textbook-driven curriculum (see Long, 2015 for commentary on such systematic teaching of *forms* as an unnatural way of imposing an order of acquisition on learners). Disgruntled with the lack of autonomy this approach gave both me as instructor and the students, I created a course around board games designed to lead learners on a journey from "noob" to "expert" in terms of 1) their English language use, 2) their board game literacy, and 3) their ability to become self-directed learners within the classroom. Previously, I have introduced the rationale for the teaching framework (York & deHaan, 2018), as well as a step-by-step walkthrough of the class in action (York, 2019c). The effectiveness of the method in terms of the goal of improving speaking skills has yet to be assessed, which is the impetus for writing this paper.

In this paper, I take a microgenetic view of how learners engage with certain scaffolding activities of the framework and how these activities positively affect learners' language development. The three activities used here are self-transcription, error-correction of transcribed data, and watching native speakers play the same game. The aim of these activities is to promote learners to "notice" mistakes in their production and improve their L2 as a standalone activity outside of gameplay as part of explicit language-focus classes (Schmidt, 1993). Activities were included based on learner feedback, teacher observations and reflection, and reference to SLA theory and research results which suggest that gameplay alone does not provide rigorous enough support for successful L2 acquisition to occur (e.g. Miller & Hegelheimer, 2006). That is, activities featuring explicit grammar instruction and text analysis in addition to gameplay are considered beneficial to learning.

The activities are completed between a first and second play session, therefore, the concept of *task repetition* is central to the framework. A brief overview follows. Each number of the list indicates a single lesson.

1. **Learn** to play the chosen game.
2. **Play** the game, record spoken audio and transcribe the audio as a homework activity.
3. **Analyze** transcriptions and watch native English speakers play the same game, noticing any useful expressions.
4. **Replay** the game, record and transcribe.
5. **Reanalyze** gameplay, compare gameplay sessions for changes in output, create a report about the game (good points, bad points, English usage, etc.).

The goal of this paper, then, is to assess whether these pedagogical interventions aided low-level learners' L2 development in a game-based teaching context, focusing on how self-transcription and watching native speakers play the same game at the post-gameplay stage influenced oral output performance during task repetition. Learner output is assessed in terms of complexity, accuracy and fluency. A description of these dimensions is provided in the methodology section.

Literature Review

Games and language teaching

There is currently a surge of research investigating the affordances of games for language learning. However, there is a trend for experimental, pilot studies around games that are carried out in laboratory or experimental contexts (Hung, Hwang, Chu, & Wang, 2018). The term digital game-based language learning (DGBLL) appears frequently in CALL literature, which I have argued emphasises technology over pedagogy and learning over teaching (York, 2019c), promoting games as "tutor" which replace a teacher rather than a tool that teachers may utilize for teaching (Levy, 1997; Reinhardt & Thorne, 2016). As a concrete example, MMOs are the core game type explored in DGBLL (Peterson, 2016). However, MMO studies are generally concerned with extracurricular or informal learning contexts which do not translate into teaching practices (Newgarden & Zheng, 2016). Examples of MMO use in classroom instruction, therefore, remains an unexplored topic.

Where classroom-based research with games has been conducted, there is an overfocus on vocabulary learning through gameplay (Enayat & Haghighatpasand, 2017; Hitosugi & Schmidt, & Hayashi, 2014; Shintaku, 2016; Zhou, 2016), which Zhou criticises as being unhelpful in supporting classroom instruction: "game-based, exploratory vocabulary learning research supports game design, not classroom instruction" (2016, p.4). Another typical thread in the literature is in assessing learner perceptions of learning with games, which often are presented without considering any learning gains (Allen, Crossley, Snow, & McNamara, 2014; Peterson, 2012; Bolliger, Mills, White, & Kohyama, 2015). One common finding related to instruction, however, is that teacher-generated materials are key to supporting learners language development (Miller & Hegelheimer 2006, Ranalli, 2008; Anderson, Reynolds, Yeh, & Huang, 2008; see also deHaan, 2019 for a critique). Materials may be implemented to support learning at the pre-play, during-play or post-play stages (Sykes & Reinhardt, 2013; York, 2019c). That is, the current body of literature suggests that gameplay should be supported with non-gameplay activities in order to allow learners to prepare for and reflect on their learning. This approach to teaching with games fits into a psycholinguistic orientation to SLA and reflects a TBLT approach to language development (see Sykes, 2014).

The methodology used in this study was informed by a TBLT perspective to SLA, and thus requires further explanation.

TBLT and board games

For low-level, monolingual EFL learners in large classroom contexts, I have argued elsewhere that the affordances of face-to-face tabletop games may provide greater benefits than digital games that typically appear in CALL and SLA research (York & deHaan, 2017, 2018; York, deHaan & Hourdequin, 2019). Another benefit of tabletop games is their ease of implementation in language classrooms due to teacher familiarity (Jones, 2019) and other logistical considerations such as cost and running costs (York, 2019c). Additionally, there are numerous parallels between a task-based language teaching (TBLT) approach to SLA (the dominant approach to language teaching in classroom contexts) and the progression of a tabletop gaming session:

1. Play follows a typical cycle similar to the pre-, during- and post-task cycle of TBLT,
2. Tabletop games afford immediate, face-to-face, social engagement, thus allowing for a focus specifically on oral communication,
3. Board games offer learners the opportunity to engage in authentic, goal-oriented activity (see Franciosi, 2011 also for a review of the links between game design and the core theoretical underpinnings of TBLT).

In sum, previous studies have highlighted the acceptability of games as a teaching tool (York & deHaan, 2018). There are however very few high-resolution, empirical reports of how games aided learners' language development. Additionally, there are few studies on how a pedagogical intervention may aid learning (deHaan, 2020). Studies thus far have focused on the affective affordances of games or vocabulary acquisition through gameplay, in other words, the use of games as content in extracurricular contexts (deHaan, 2019; Poole & Clarke-Midura, 2020).

Task repetition

In a study by Bygate (1996) task repetition was linked to improvements in output complexity and smaller gains in accuracy, however, in a follow up (2001) results suggested an improvement in fluency and complexity. Bui, Ahmadian, and Hunter (2019) also found that regardless of the length of time between a production task and its repetition, task repetition aided fluency and structural complexity. Output complexity was also improved during subsequent task performances in studies by Hawkes (2011) and Fukuta (2015). In a study by Date (2013) task repetition led to an improvement in accuracy. Finally, in a study by York (2019b) it was not complexity but accuracy that was affected by task familiarity, echoing Date's findings. In these studies then, regardless of whether complexity or accuracy improved in repeated task performance, as both of these dimensions are considered to belong to the "form-focused" dimension of performance, the above findings suggest that task repetition is effective at shifting learner attention from meaning to form as the cognitive demands of the task are reduced, possibly due to improved familiarity and thus lower cognitive demands.

Conversely, in recent studies by Wang (2014) and Sample and Michel (2014), output fluency improved in subsequent task performance. Wang collected data from 77 undergraduates who repeated a video-based narrative task where results showed an improvement in all three dimensions. Sample and Michel, learners completed the same spot-the-difference task a total of three times. During the first repetition, fluency scores improved at the expense of complexity and accuracy. However, the trade-off disappeared in the second repetition leading the authors to conclude that familiarity with task content allows learners to progressively focus on all three CAF dimensions.

In summary, then, whilst a number of studies have shown that task repetition allows learners to increase their attention to form (complexity and accuracy), task repetition is considered a robust method of improving output performance along any of the CAF dimensions.

Self-transcription

Self-transcription can promote metalinguistic awareness, or what has been called "noticing" from an interactionist perspective to SLA (Mennim, 2012; So, 2015; Salas, 2015). Studies have shown self-transcription to promote syntactic, pragmatic, lexical and fluency improvements (deHaan, Johnson, Yoshimura & Kondo, 2012; Lynch, 2007; Cowie, 2018). Through this post-play activity then, students are being promoted to be more purposeful in their learning. And, with the inclusion of a group analysis session, a ZPD of expertise may be created amongst the learners as they notice their mistakes, others' mistakes, and are self-determined in regards to the errors that they focus on. Additionally, from a TBLT approach, self-transcription may be considered a "focus on form" task raising learner consciousness around their mistakes and aid in accuracy development (Lynch, 2001).

In pre-smartphone days, transcription was difficult to perform as a classroom activity, however, due to the ubiquitous nature of smartphones nowadays, recording, sharing and transcribing audio is relatively easy (Cowie, 2018). So (2015) provides a succinct summary of studies which utilized self-transcription including pedagogical procedure and results (including Lynch, 2001, 2007; Mennim, 2012; Stillwell et al., 2010). A typical procedure follows:

1. Complete a spoken mono- or dialogic activity (typically roleplay or presentation)
2. Transcribe a section of it. (90 seconds to 5 minutes based on length of activity)
3. Assess performance
4. Teacher corrects errors
5. Repeat the process
6. [Optional] Compare the two transcriptions

According to the above procedure then, upon completing transcriptions, learners often assess their performance by identifying errors. Lynch (2001) and Mennim (2007) employed pairwork at the transcription analysis phase which accounted for a considerable increase in error identification (50% of errors identified by peers in Lynch, 2001), whereas Salas (2015) had learners analyse transcriptions individually and found that only 25% of errors were noticed. In the present study, the whole group of participants shared their transcriptions of gameplay and analysed it together.

Transcription was included in this pedagogical intervention as it was discovered through teacher observations of preliminary versions of the framework that gameplay presented considerably high cognitive demands on learners, leaving them with limited cognitive capacity available to focus on form. Thus, based on the findings of other studies of self-transcription, post-play participants analysis of their own performance in accordance with watching native speakers play was considered a useful tool in promoting the noticing of errors and improving task performance in subsequent gameplay sessions (Stillwell, et al. 2010; deHaan, Johnson, Yoshimura & Kondo, 2012). This paper investigates the impact of these post-play activities.

Transcription and task performance

Skehan and Foster (1997) found that self-transcription led to a significant accuracy effect in repeated performances (see Skehan, 2001 also). Subsequently, Foster and Skehan (2013) conducted a study comparing the performances of a post-task self-transcription group with a control group which did not perform any post-task activity. Of note is that task-repetition was not conducted. The experiment was designed to explore whether foreknowledge of a post-task self-transcription task would be enough to

promote improved performance. Results suggested that, of two different task types (decision making and narrative), 1) foreknowledge of the post-task self-transcription task affected output accuracy for both tasks, 2) complexity and fluency were improved for the decision making task type.

Lynch (2007) compared the benefits of post-task transcription on learners' speaking performance. One class used teacher-generated transcripts, the other used learner-generated transcripts. Analysis of two subsequent performances of the same initial task revealed that self-transcription was more effective at promoting accurate language use than using teacher-generated transcripts. However, it has been noted in the literature that learners' often require additional instructor support in noticing errors in their transcriptions (Lynch, 2001). For example, Salas (2015) found that learners only recognised 25% of their errors in a three minute transcription. Indeed, Abadikhah and Valipour (2018) conducted a study which compared two experimental groups. One group engaged in self-transcription only, the other engaged in self-transcription which was supplemented with instruction. Results suggested that instruction aided learner accuracy in two of five dimensions measured.

In summary then, self-transcription has been linked to improvements in all three dimensions of performance. Due to the ease of promoting learners to find and correct errors in their production, accuracy appears to be most positively affected in subsequent performances. Accuracy can also be improved with additional instruction, promoting further noticing of errors. However, as Mennim (2007) notes, there is no guarantee that forms noticed in self-transcription activities would re-emerge as output during subsequent performances (p. 272). The current study investigates this point as part of the research questions which are provided below.

Research questions

1. Did the pedagogical intervention help participants' oral performance improve during the second gameplay session?
2. What errors did participants correct?
3. What expressions did participants notice in their video viewing sessions?
4. Did participants' post-play error corrections and expressions from video viewing appear in the replay session?

Methodology

Context

The current study may be considered a single-case study where data was collected from a single group of students in an intact class at a private science and technology university in Japan. The single game-playing group here is considered the case and individuals within the group are considered as embedded subunits of the case (see Yin, 2018). The class of 28 students met once a week. The goal of the course was to improve students' communicative English ability and the researcher of this study acted as the instructor.

Participants

The participants were recruited through convenience sampling. This study focuses on a single group of nine students. They were all male, Japanese, and had an age range of 18-20. All participants studied computer science and had reasonably similar backgrounds in terms of English education. None of them had taken a standardized English test (such as TOEIC). Their class was the third-highest of six classes based on a placement test, and they may be described as low-proficiency English speakers.

Procedure

This study was framed from a TBLT perspective to second language development where gameplay is supported with pre- and post-play activities. A simple overview of the class progression (task cycle) is provided in Figure 1 where each box represents a single lesson.

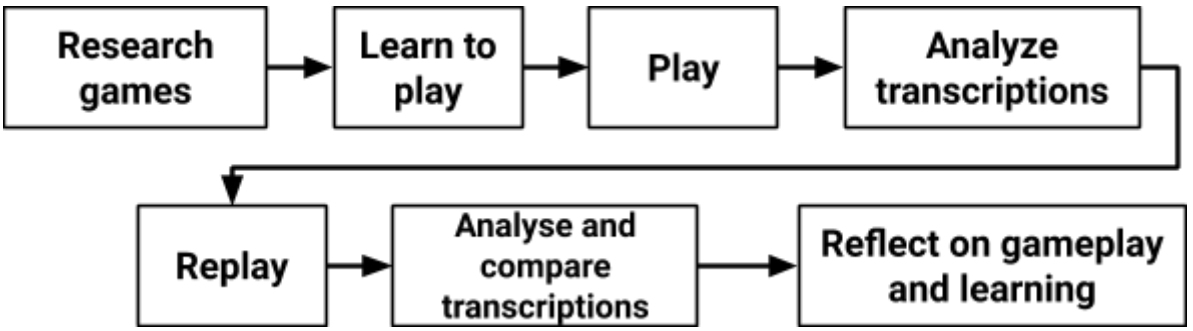


Figure 1 A graphical representation of a task cycle.

Lessons were 100 minutes in length, and 14 lessons were conducted per semester. Data collected as part of this study was collected from all stages of the intervention except for the first class where learners researched games to play. Additionally, data were collected at the start of the second semester, meaning that the participants had already completed the task cycle two times previously. The first two cycles featured extensive teacher mediation in the form of a workbook which contained grammar guides, questions regarding game rules, transcriptions of play sessions, and video analysis commentary and support (York, 2019a). In the second semester the workbook features less predefined mediation, but space for learners to fill in themselves. This was a deliberate part of the syllabus design whereas students completed the teacher-led cycles of gameplay and analysis, they would learn how to complete the activities themselves. Hence, the third and fourth cycles were designed to be carried out with an emphasis on student-centred exploration. Teacher mediation still played an important role, but was relegated to retroactive oral feedback rather than through predefined materials. See Appendix 1 for a comparison of teacher-led and student-centred activities from the workbook.

Student activities and teacher mediation are provided in Table 1. For a more detailed exploration of student and teacher activities, see York (2019c).

Table 1 *Student and teacher activities in the five weeks used to gather data for this study.*

Wk	Class activities	Extracurricular activities	Teacher mediation
1	**Pre-play: Learn** ● Read the rulebook ● Watch rule explanation videos ● Write questions about the rules ● Test play and consider important words/phrases needed to play	Review notes and prepare for the following play session	● Assist student groups in understanding game rules. ● Promote learners to complete the relevant sections of the workbook. ● Emphasise the importance of preparing for play by brainstorming potentially useful expressions.
2	**Play** ● Play and record the game with smartphones ● Transcribe gameplay audio	Complete the transcription of gameplay audio	● Provide in-time feedback on errors. ● Make notes on common errors for students to refer to after playing.
3	**Analyze (see Appendix 2)** ● Find errors in the transcription ● Translate Japanese utterances to English ● Watch online gameplay videos and make notes on useful expressions	Review notes and prepare for the following play session	● Check students transcriptions for errors. ● Promote students to look up specific grammar points. ● Help students understand what is being said in the videos.
4	**Replay** ● Play the game again ● Transcribe gameplay audio		● Instruct students to drill and practice useful expressions before playing ● Assess students' performances
5	**Reanalyze** ● Find errors in the second transcription ● Analyze how many times they used expressions from the first analysis session.	N/A	● Ask students to consider why they did or did not use specific expressions

*Note: Highlighted areas are where data were collected in this study.

In summary then, students completed three different post-play activities:

1. Self-correcting English errors from a transcription of their gameplay session
2. Translating Japanese utterances into English (from the same transcription)
3. Collecting useful expressions from watching native speakers play the same game.

The researcher also created a transcription of both play sessions based on audio and video recordings, giving an accurate transcription to use in order to answer RQ1: Did participants' oral performance improve during the second gameplay session?

Game used

The game selected by participants in this paper is called *Mafia de Cuba* (des Pallières & Lamy, 2015). It is an asymmetric hidden-role game in which a mafia boss known as the "Godfather" must recover stolen diamonds from thieves hidden among the other players. First, the Godfather removes a certain number of diamonds from a cigar box which contains both diamonds and character tokens. Then, the

box is then passed clockwise around the group where players may take either diamonds (becoming a thief) or a character token (becoming that character). Players are either on the side of the godfather (known as the "loyal henchmen") and help him recover the missing diamonds or are thieves and try to persuade the godfather to accuse one of his loyal henchmen of stealing diamonds. If the godfather can correctly identify all thieves and recover all diamonds, he and his loyal henchmen win. If the godfather incorrectly accuses two of his loyal henchmen, he loses and the thieves win. A photo of the group playing this game is available in Figure 2. Of note in this photo are participants' smartphones on the table recording gameplay audio.

Figure 2 *Participants playing Mafia de Cuba.*

Self-Transcription (student activity)

Self-transcription was carried out through the use of participants' smartphones as in Cowie (2018). Audio was divided evenly between the participants and transcribed at the end of gameplay classes (2 and 4) and for homework before analysis classes (3 and 5). The accompanying workbook for the course featured specific pages for participants to complete their self-transcriptions (see Figure 3). As the transcription activity was such an integral part of the class, the self-transcription homework was assigned a certain number of points towards their overall grade.

For clarification, the student-created self-transcriptions were not used as a data source to answer RQ1 (Did participants' oral performance improve during the second gameplay session?), only RQ2, 3, and 4. In other words, (as shall be made clear in the Results section below) student created self-transcriptions were not complete and not accurate enough to be used to answer RQ1. Thus, the researcher also created accurate transcriptions of both play sessions.

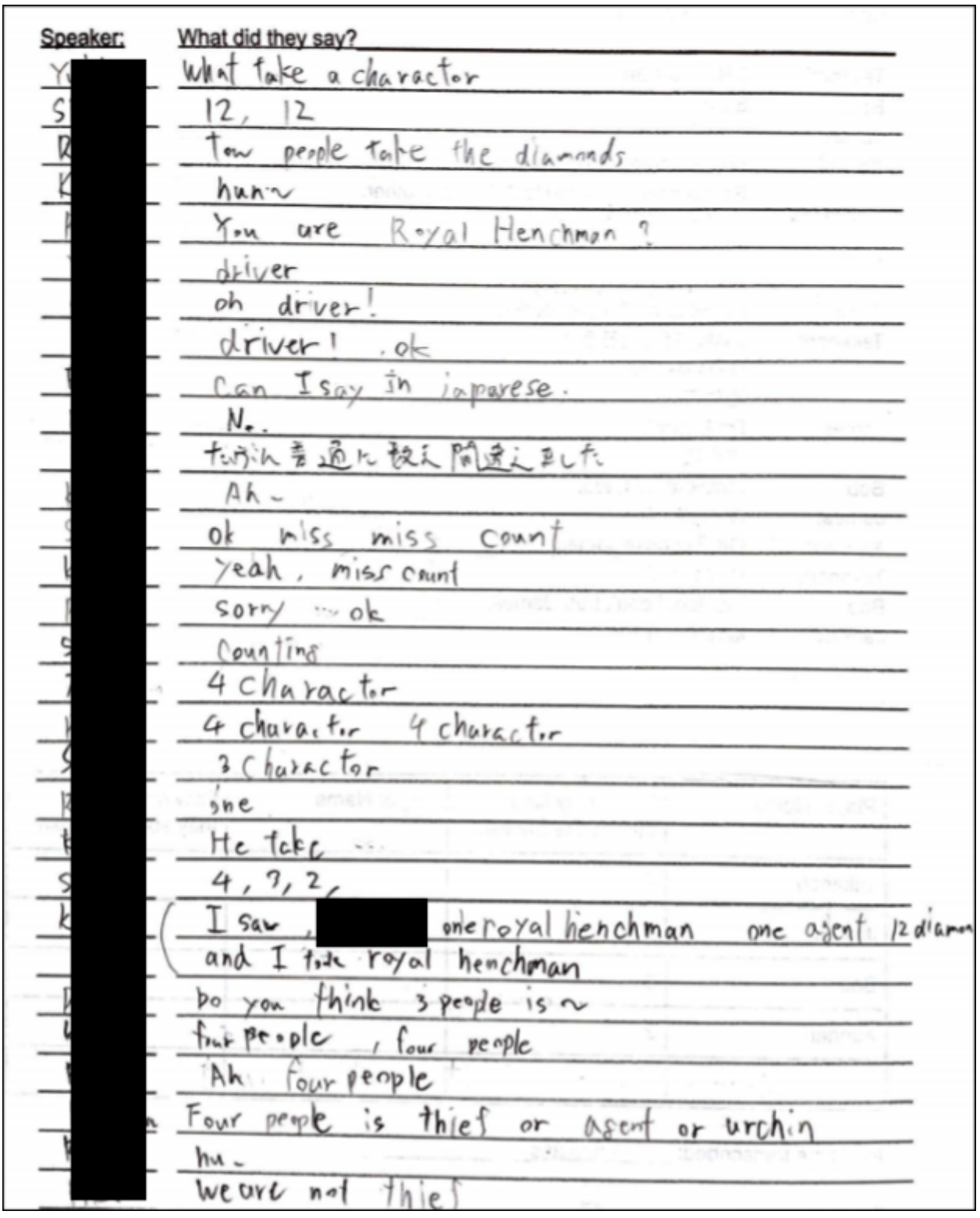


Figure 3 *An example of self-transcribed audio (names redacted).*

Native speaker video analysis

During the Analyze and Reanalyze classes, students were required to watch online gameplay videos of English speakers playing the same game. Participants were instructed to search for videos with subtitles by using YouTube's filters. However, in lieu of not finding a suitable video with subtitles, they were also prompted to consult with me, the instructor, regarding any areas of the video that they wished to analyze in further detail. For the group focused on in this paper, they watched the following

video: Mafia de Cuba by Board and Chill[1]. See Appendix 2 for the page of the workbook that participants were required to complete for this activity.

Data analysis

CAF Measures

Skehan (1998) proposed the CAF (complexity, accuracy and fluency) model as a way for researchers to understand learner proficiency through the collection and analysis of quantitative data. Fluency is a measure linked to learner focus on meaning, whereas complexity and accuracy are considered measures of learner attention to form (Fukuta, 2015). Since its inception, a plethora of measures have been introduced for assessing learner performance.

Some key measures for each of three dimensions are as follows:

Complexity
- Number of different words spoken (types)
- Token type ratio
- Number of syllables per utterance
- Measure of Textual Lexical Density (MTLD)

Accuracy
- Error-free clauses

Fluency
- Temporal fluency
    - Rate of speaking
    - Length of fluent runs before a pause
- Vocal fluency
    - Number of false starts
    - Reformulations

The scope of this paper does not allow for a detailed overview of all measures, however, interested readers are encouraged to read Jackson and Suethanapornkul (2013) who categorize 84 measures for L2 production.

Fluency is a measure linked to learner focus on meaning, whereas complexity and accuracy are considered measures of learner attention to form.

In order to understand how the post-play transcription and video watching activities affected learner performance, oral complexity accuracy and fluency was assessed quantitatively. As mentioned above, there are an abundance of different measures available for assessing learner output. In this study, the following measures were chosen based on their appropriateness as measures of oral performance, their perceived reliability, and based on adoption rates in other, relevant studies.

Complexity

Complexity was assessed with the following two measures.

- Syllables per utterance (Yuan & Ellis, 2003; Lintunen, & Mäkilä, 2014),
- Measure of textual lexical diversity (MTLD) (Koizumi, 2012).

Initially, clauses per *Analysis of Speech Unit* (AS-Unit) was chosen as a suitable measure however, was later rejected based on initial transcription data where there were very low numbers of utterances with

[1] https://www.youtube.com/watch?v=GgOr73d0fHI

more than once clause per AS-Unit. York (2019b), conducted in the same context, featured the same finding.

Although space limitations hinder a detailed discussion on the different segmentation units available for measuring oral performance, suffice it to say that an AS-Unit is defined as "a single speaker's utterance consisting of an independent clause or sub-clausal unit, together with any subordinate clause(s) associated with either" (Foster et al., 2000 p. 365). This unit is therefore designed with consideration of the specific features of speech rather than written output. For written output the T-unit or C-unit are often used equivalents[2].

Accuracy

Accuracy was measured as the number of error-free AS-Units (Sample and Michel, 2014; Amiryousefi, 2017). Errors were also classified into three subcategories based on lexical, morphological and syntactic error types.

Fluency

Fluency was measured by the temporal fluency measure (as opposed to vocal measures) as the total number of syllables per minute (Kormos & Trebits, 2012; York, 2019b). This measure was chosen over vocal fluency measures due to ease of analysis.

Statistical tests

Due to the low samples number, the current study presents descriptive statistics only, apart from a single linear stepwise regression analysis for students accurate output. In addition, quantitative data (collected from participants' written work) is triangulated with qualitative data in order to provide further evidence for any findings.

Researcher transcription considerations

In order to ascertain whether self-transcription and video viewing sessions promoted improved oral performance in the repeated gameplay task, **audio data from both sessions were collected and transcribed by the researcher** (that is, in addition to participants' self-transcriptions). An example can be seen in Table 2. Accuracy codes for the different error types are as follows:

- Y -- error-free utterance.
- L -- utterance contains a lexical error.
- M -- utterance contains a morphological error.
- S -- utterance contains a syntactic error.

[2] For a discussion on the various segmentation units available for measuring written and spoken output see Foster, Tonkyn, and Wigglesworth (2000).

Table 2 *Example of transcribed audio*

Utterance Number	Participant	Utterance	Syllables	Accuracy codes
32	Ku	I look in the box... there is... there are twelve diamonds and one loyal henchmen and one agent	22	m l
33	Ku	And I took loyal henchmen	12	l
34	Ku	These three players loyal henchmen or driver and I am loyal henchmen	18	l m
35	Sh	Loyal henchmen	4	m
36	Ju	There are four character.	6	m
37	Ju	When I took a character, I saw four characters	13	y
38	Ju	Loyal henchmen, henchmen, driver, Agent	10	m
39	Kob	He said lie	3	l m
40	Sh	Liar.	2	
41	End	Four, three	2	
42	Ju	When I take て	3	m

Audio data from both sessions were collected and transcribed by the researcher

Results

In this section the transcription and analysis work completed by the participants is presented alongside the authors own transcription and analysis.

The gameplay sessions are coded accordingly:

- Gameplay 1: Initial gameplay in the Play class (class 2)
- Gameplay 2: Gameplay in the Replay class (class 4)

Features of participants self-transcription

Due to the limited space provided for transcription in the workbook, **a full transcription of gameplay was not produced by participants**. This is not necessarily a problem when one considers the repetitive nature of their communication. The collation of their individually transcribed sections of the overall gameplay audio may be adequate enough data for assessing what errors were common. A detailed analysis of typical self-transcription features follows.

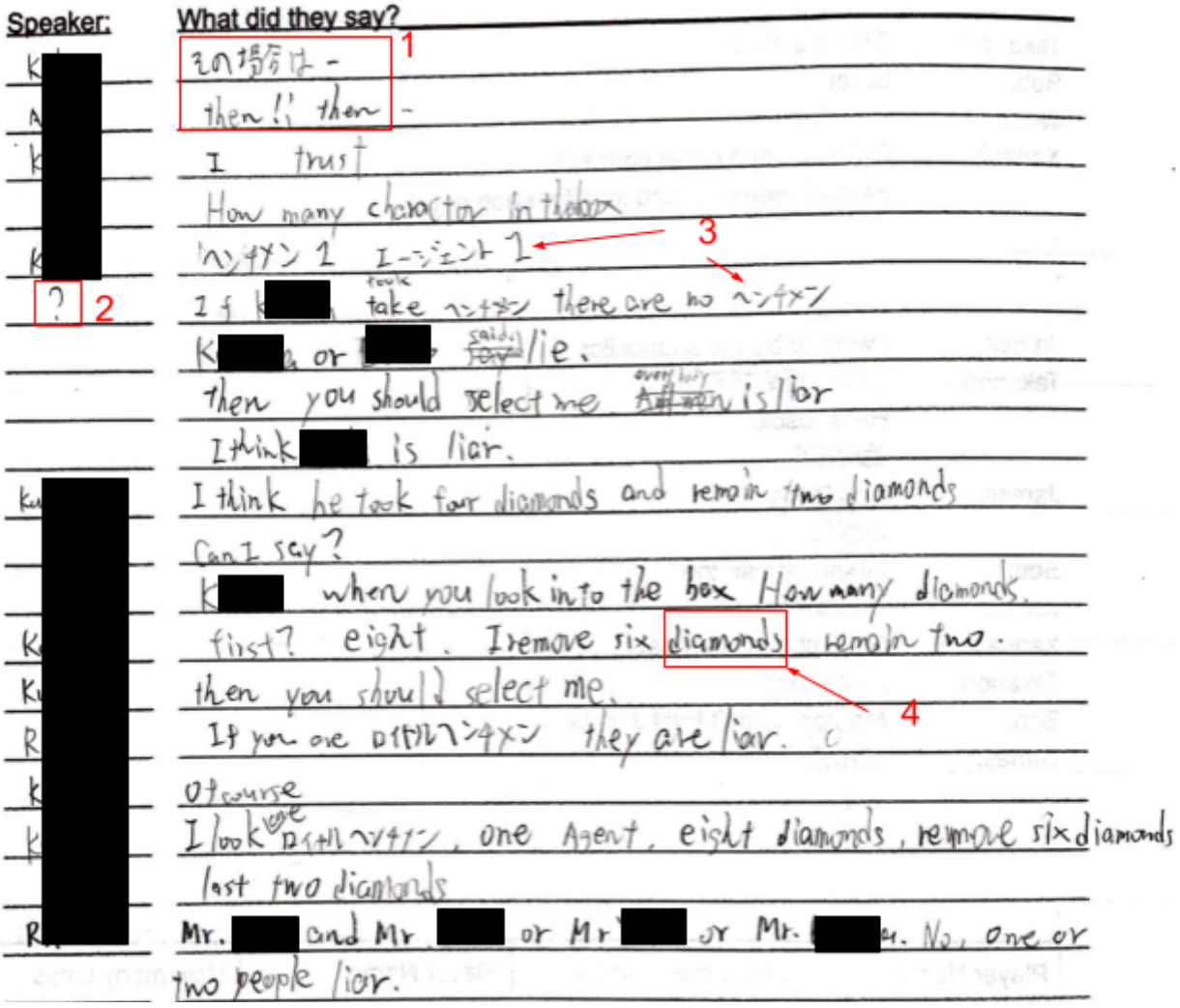


Figure 4 *An example of a participant's self-transcription from Gameplay 1.*

Figure 4 is an excerpt of a participant's transcription which was selected based due to its typicality in regards to the common features of participant transcriptions:

1. The transcription of both English and Japanese utterances.

This was a prescribed activity. In other words, participants were instructed to transcribe the audio verbatim, including any Japanese utterances. Both English and Japanese utterances from the data are used in the post-play analysis activities.

2. Using question marks to represent an unknown speaker.

One limitation of recording gameplay audio with smartphones is the fidelity of captured audio. If students knock the table or move objects it can obscure participant utterances. Additionally, with a large group of students (as in this case study) it can be hard to tell who is speaking, especially if a group is conducting gameplay for the first time and they are not familiar with each other's voices. A video recording may help alleviate such issues, but recording and distributing video data between participants poses considerable logistical issues for classroom implementation. In other words, audio can be captured very easily and on an individual basis with participants' smartphones; video requires more careful planning and set up, the use additional equipment (tripods), and, most importantly requires additional, valuable class time.

3. Use of *katakana* to transcribe English words.

Participants often wrote out English words in the Japanese syllabary *Katakana*. Although I do not have empirical data on why they did this, the following propositions may be made. 1) Although participants

read the game rulebook as a pre-play activity, they do not have access to it during their transcription homework and thus spelling of words that are specific to the game are unknown at the time of transcription, or, 2) transcription is an activity that requires the extensive use of short-term memory. The cognitive load of remembering what was said may take priority over the correct spelling of words. In the above example for instance, the word "henchmen" is written out in katakana as ヘンチメン (henchimen), a word that does not exist in this form in Japanese, thus used as scaffold to the English spoken during gameplay.

4. Inaccuracies in transcribed data.

The final, and perhaps most common feature was transcription inaccuracies. As seen in the above excerpt (highlighted as No. 4), the participant has inaccurately transcribed the word "diamonds." That is, listening to the audio myself, it is clear that the participant who produced this utterance said "six diamond," thus the singular version of the noun and a morphological error. It is unknown whether such inaccuracies were a deliberate attempt to make gameplay speech seem more accurate than it really is, or whether the participants are instead transcribing player intention rather than the actual produced utterances. Stillwell et al. (2010) explored transcription errors systematically in a context similar to the present study (at a Japanese university and over multiple instances) and found that on average, learners produced 19 transcription errors on their first attempt, but only 13 on subsequent sessions. Inaccuracies are thus a common feature of learner transcriptions, however, they are considered a positive indication of task performance (Stillwell et al., 2010).

Did participants' oral performance improve over repeated gameplay sessions?

This section presents results of statistical analyses conducted on the quantitative data generated from participants' oral performances for both the initial gameplay and repeated gameplay sessions. Using the CAF measures defined above, participant performance improvements are presented systematically from complexity, accuracy and finally fluency dimensions.

Output complexity

Syllables per utterance

Inspection of descriptive statistics for the syllables per utterance measure reveals that on average, participants produced 4.78 syllables per utterance during gameplay 1, with only minimal changes in gameplay 2 (Table 3).

Table 3 *Descriptive statistics for the number of syllables per utterance measure for each participant*

Participant	Gameplay 1			Gameplay 2		
	Number of syllabus	Number of utterances	Syllables per utterance	Number of syllabus	Number of utterances	Syllables per utterance
A	16	3	5.33	54	10	5.40
El	146	28	5.21	113	16	7.06
En	185	42	4.40	46	10	4.60
Ju	159	27	5.89	33	10	3.30
Ko	212	58	3.66	131	32	4.09
Ku	434	77	5.64	307	56	5.48
O	31	8	3.88	195	43	4.53
R	867	154	5.63	331	62	5.34
S	245	73	3.36	89	23	3.87
M	**255**	**52.22**	**4.78**	**144.33**	**29.11**	**4.85**
SD	**260.25**	**46.29**	**0.96**	**110.86**	**20.34**	**1.11**

Focusing solely on the mean scores for the number of syllables per utterance participants produced in gameplay 1 and 2 shows a minimal change between the two gameplay sessions (4.78 in gameplay 1 and 4.85 in gameplay 2). However, this does not capture the variance in participant output as they completed both gameplay tasks. The number of utterances produced by each participant during gameplay 1 is incredibly varied (SD = 46.2). Participant A produced only 3 utterances, whereas Participant R produced 154. In gameplay 2 however, the number of utterances produced by each participant are less varied (SD = 29.11). Participant R still produced the most (62) but there are three participants that share the lowest number of utterances (10). The reason for the more uniform number of utterances in gameplay 2 could be due to participants' familiarity with the game, but also due to which role they were playing between games. That is, there is an expectation for certain roles to speak more than others during this game which may have had a strong influence on level of participation.

MTLD

There was a problem encountered when trying to assess the MTLD score for each participant. As can be seen in Table 4, it was not possible to record scores for a number of participants due to the small sample size recorded for those participants. The MTLD test requires a minimum number of tokens (words) in order to run, and in five cases, this minimum value was not exceeded. That is, a number of the participants (particularly Participant A) produced very few utterances during gameplay. As such, participants that did not produce enough output to be analyzed had their corresponding scores redacted. For gameplay 1 this is Participants En and Ju, for gameplay 2 this is Participant O.

Inspection of descriptive statistics for the MTLD measure reveals that on average, participants scored 26.54 (SD=10.08) for gameplay 1 and 23.68 (SD=9.23) for gameplay 2. It should be noted, however, that there was a considerable difference between scores for each participant (resulting in the large SD values). Results for this measure therefore reveal no difference in output complexity in terms of MTLD.

Table 4 *Descriptive statistics for the MTLD measure for each participant*

Participant	Gameplay 1	Gameplay 2
A	--	--
El	34.26	32.47
En	~~18.94~~	--
Ju	~~32.91~~	--
Ko	17.81	18.35
Ku	37.93	20.06
O	--	~~14.17~~
R	27.93	34.29
S	14.75	13.22
M	**26.54**	**23.68**
SD	**10.08**	**9.23**

Output accuracy

Inspection of descriptive statistics for the error-free utterances measure reveals that on average, 37.57% (SD=0.2) of participant utterances were error-free during gameplay 1. This score was improved markedly during gameplay 2 (M=58.06%, SD=0.12) (Table 5). Of the 9 participants, only two had lower accuracy scores in gameplay 2 (Participants A and O), but this may be a result of them producing more utterances during gameplay 2. The opposite could be said of other participants however, as R, S and Ku all produced less utterances, but of higher accuracy. Perhaps during gameplay 2 then, these participants took their time and produced less utterances, focusing on accurate output.

Table 5 *Descriptive statistics for the percentage of error-free utterances for each participant*

Participant	Gameplay 1		Gameplay 2	
	Number of utterances	Error-free utterances (%)	Number of utterances	Error-free utterances (%)
A	3	50.00%	10	44.44%
El	28	20.83%	16	71.43%
En	42	25.71%	10	50.00%
Ju	27	20.83%	10	60.00%
Ko	58	21.74%	32	42.11%
Ku	77	34.62%	56	61.36%
O	8	75.00%	43	50.00%
R	154	29.77%	62	68.18%
S	73	59.65%	23	75.00%
M	**52.22**	**37.57%**	**29.11**	**58.06%**
SD	**46.29**	**0.2**	**20.34**	**0.12**

A stepwise linear regression analysis in R (R Core Team, 2017) was run to determine if participants' accuracy was affected by the number of utterances they made and gameplay session[3]. To ensure there were no violations of assumptions of normality, linearity, multicollinearity, and homoscedasticity, preliminary analysis was first conducted and no violations were identified. For both models, output accuracy (percentage of error-free utterances) was used as the dependent variable. Then, the total number of utterances was entered into the model followed by the gameplay session as a variable. An ANOVA was then used to determine if the second gameplay session contributed significantly to participants' output accuracy.

The second model with the addition of gameplay session explains significantly more variance in the percentage of error-free utterances than the null model ($X2\ (1)= 5.21$, $p=.037$) and is significant ($F(2,15)=3.63$, $p<.05$) with an R^2 of .326 (Table 6). The addition of a second gameplay session accounts for approximately 23% more variance than the null model. In other words, participants were able to improve their output accuracy between gameplay sessions, which can be attributed to the work they completed as part of the post-task activities.

[3] Which, in turn implies that the post-task activities were effective in promoting participants' consciousness regarding their output accuracy.

Table 6 *Results of linear stepwise regression on participants accuracy.*

	Dependent variable:	
	Output accuracy	
	(1)	(2)
Number of utterances	-0.002	-0.001
	(0.001)	(0.001)
Gameplay 2		0.188**
		(0.083)
Constant	0.542***	0.413***
	(0.067)	(0.082)
Observations	18	18
R^2	0.092	0.326
Adjusted R^2	0.035	0.236
Residual Std. Error	0.186 (df = 16)	0.166 (df = 15)
F Statistic	1.624 (df = 1; 16)	3.632* (df = 2; 15)
Note:		*p<0.1; **p<0.05; ***p<0.01

For error types, descriptive statistics revealed that participants made the most morphological errors in both gameplay sessions, followed closely by lexical errors (Table 7). Syntactic errors were rarely produced in either session. The most common lexical error was the omission of the articles “a” or “the” before nouns. The most common morphological error was in use of the singular form where a plural form was required, e.g. saying “three diamond” instead of “three diamonds.” For syntactic errors, the most common error type was for word order in question formation such as this utterance from Participant R: “And, Mr. Kuroda, you saw what kind of character token?*”

Table 7 *Descriptive statistics for erroneous utterance types.*

Participant	Gameplay 1			Gameplay 2		
	Lexical	Morph.	Syntactic	Lexical	Morph.	Syntactic
A	25.19%	39.69%	5.34%	15.91%	13.64%	2.27%
El	30.43%	43.48%	4.35%	31.58%	26.32%	0.00%
En	34.62%	26.92%	3.85%	22.73%	15.91%	0.00%
Ju	25.00%	50.00%	4.17%	20.00%	20.00%	0.00%
Ko	17.54%	22.81%	0.00%	6.25%	18.75%	0.00%
Ku	0.00%	25.00%	0.00%	10.53%	34.21%	5.26%
O	29.17%	45.83%	4.17%	7.14%	21.43%	0.00%
R	42.86%	22.86%	8.57%	20.00%	30.00%	0.00%
S	25.00%	25.00%	0.00%	44.44%	11.11%	0.00%
M	**25.53%**	**33.51%**	**3.38%**	**19.84%**	**21.26%**	**0.84%**
SD	**0.119**	**0.111**	**0.029**	**0.122**	**0.076**	**0.018**

Note: Scores are derived from the total percentage of all utterances including error-free utterances.

Output fluency

Inspection of descriptive statistics for the syllables per minute measure reveals that on average, participants produced 9.43 (SD=9.07) syllables per minute in gameplay 1. This score was reduced in gameplay 2 to M=8.53 (SD=6.18) syllables per utterance in gameplay 2 (Table 8). It should be noted that participants' scores varied significantly for both games, with two players dominating the number of utterances in the first game, and to a certain extent, the second game also. A possible reason for this is explored in the discussion section below.

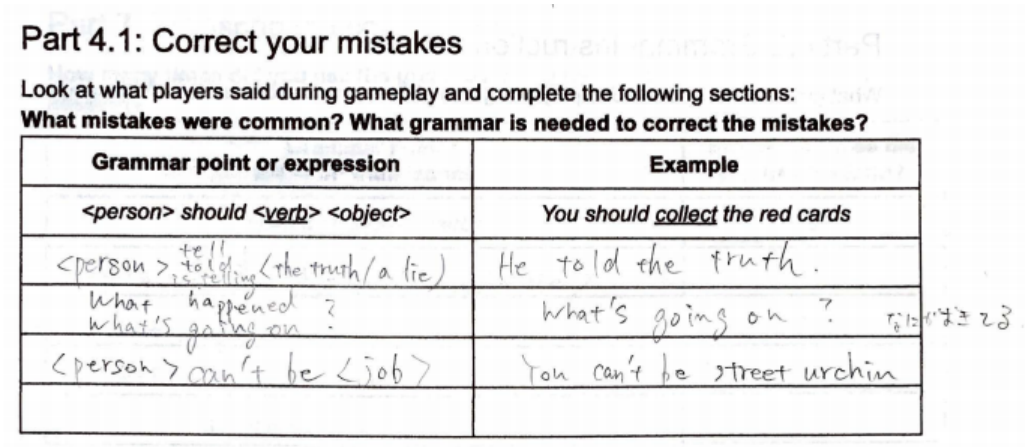
Table 8 *Descriptive statistics for the percentage of syllables per minutes for each participant*

Participant	Gameplay 1	Gameplay 2
A	0.59	3.19
El	5.40	6.68
En	6.84	2.72
Ju	5.88	1.95
Ko	7.84	7.74
Ku	16.04	18.15
O	1.15	11.53
R	32.05	19.57
S	9.06	5.26
M	9.43	8.53
SD	9.07	6.18

What errors did participants correct?

This section explores the errors that participants found in their self-transcribed performance. All of the data was thus generated from the post play activities that students completed as a group. The content of Tables 10, 11 and 12 are therefore gathered from all participants (see Figure 5). Upon completion of the task cycle, participants' workbooks were collected and scanned to generate data regarding the errors they corrected (Table 10), the Japanese they translated (Table 11), and the expressions they deemed useful from watching native speakers play on YouTube (Table 12).

Firstly, (Table 10) introduces the errors that participants corrected, It should be noted here that they adopted the formatting that was prescribed by the teacher. For example the use of less than and greater than signs (< >) to highlight a section of a sentence that can be freely changed.



Part 4.1: Correct your mistakes

Look at what players said during gameplay and complete the following sections:
What mistakes were common? What grammar is needed to correct the mistakes?

Grammar point or expression	Example
<person> should <verb> <object>	*You should collect the red cards*
<person> tell/told/is telling (the truth/a lie)	He told the truth.
What happened? What's going on.	What's going on?
<person> can't be <job>	You can't be street urchin

Figure 5 *An example of a participants workbook showing what errors they corrected*

Table 9 *Errors participants corrected from the first transcription*

Error identified	Error type
<person> is telling <the truth>/<a lie>	Morphological (tell → telling) Lexical (the truth / a lie)
What's going on?	Lexical (use of the chunk “What’s going on?”)
<person> is suspicious	Lexical (use of the word “suspicious”)
I saw...	Morphological (see → saw)
<person> told the truth	Morphological (tell → told) Lexical (say → tell)
What happened?	Morphological (happen → happened)
<person> can't be <job>	Morphological (is not → can’t be)

What is clear from Table 9 is that participants identified and corrected a number of lexical and morphological errors from their transcriptions to aid their second gameplay session. However, they have not recognised the two main causes of errors: omission of articles and use of singular rather than plural forms. This relates to the findings of Lynch (2001) and Salas (2015) in that although participants noticed errors, additional teacher mediation is still required. Lynch identified vocabulary as an area that required additional instruction, however, in the present study grammatical errors were more prominent.

The next section of the workbook asked participants to identify L1 (Japanese) utterances in their transcriptions and find equivalent English expressions for use in gameplay 2 (Figure 6). The expressions they translated are available in Table 10).

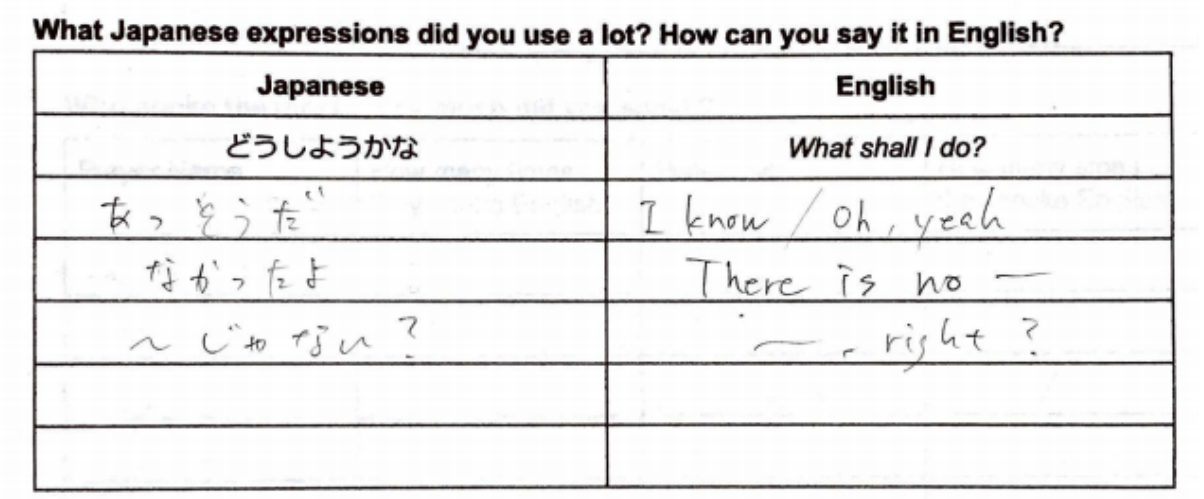


Figure 6 *An example of a participant’s workbook showing expressions translated from Japanese to English*

Table 10 *Japanese utterances translated into English*

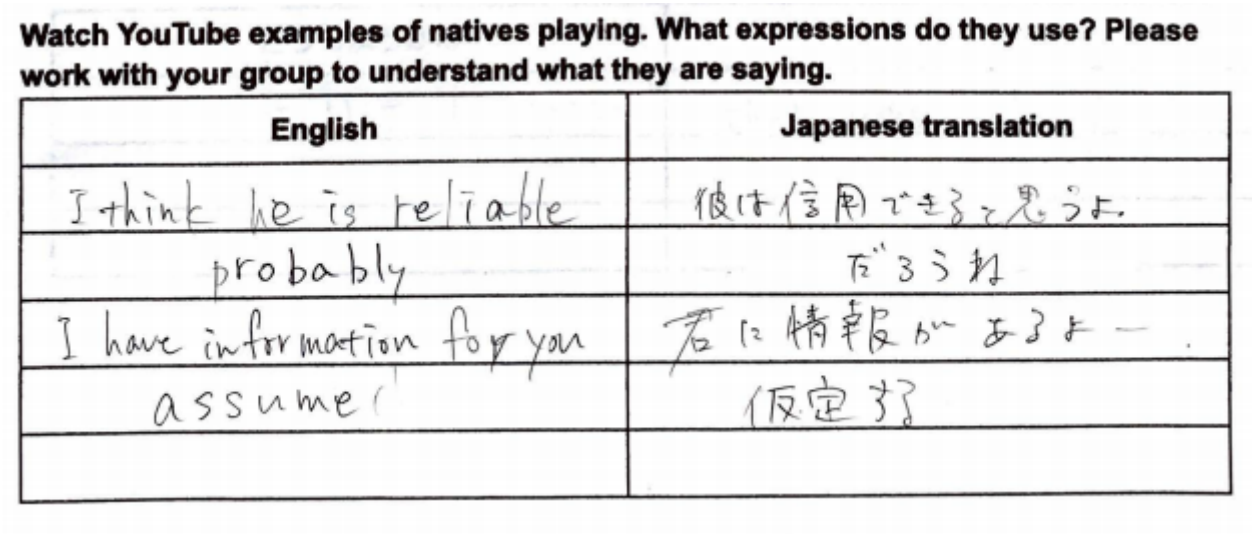
Japanese used in gameplay 1	Translated to English
あっ、そうだ (A, sou da)	I know! / Oh, yeah!
なかったよ (nakatta yo)	There is no...
～じゃない？ (~jyanai?)	..., right?

They identified three phrases, two which appeared in the transcripts and one of which was proposed by the instructor: the use of the question tag “right?” as a way to clarify information.

What expressions did participants notice in video viewing sessions?

The third section of the workbook was provided for participants to collect vocabulary or phrases that they heard native speakers use during their gameplay. Before exploring the data, it is worth mentioning here, that participants considered this activity both enjoyable and valuable. Having gained contextual experience of the game from their first play session, participants could learn how native speakers approached gameplay and were able to compare the phrases native speakers used to their own performance. The activity was thus considered valuable in gaining English cultural knowledge and linguistic resources for the second gameplay session (see York, 2019c, p.94 for more details on this topic).

Participants appeared to focus on key phrases that would aid the progress of gameplay and a number of keywords that they had not used during gameplay 1 (Figure 7 and Table 11). Whether these phrases appeared in the gameplay 2 is the subject of the next section.



Watch YouTube examples of natives playing. What expressions do they use? Please work with your group to understand what they are saying.

English	Japanese translation
I think he is reliable	彼は信用できると思うよ.
probably	だろうね
I have information for you	君に情報があるよ.
assume	仮定する

Figure 7 *An example of a students' workbook with expressions collected from YouTube video watching*

Table 11 *Phrases collected from YouTube video watching*

Noticed phrase	Japanese translation
I think he is reliable	彼は信用できると思うよ (Kare ha shinyou dekiru to omou)
probably	多分。だろうね (tabun / darou ne)
I have information for you	君に情報があるよ (Kimi ni jouhou ga aru yo)
assume	仮定する (katei suru)

Did participants' post-play error corrections and expressions from video viewing appear in the replay session?

The final research question explores whether errors and phrases collected during post-task analysis appeared in gameplay 2. Data regarding the appearance of phrases in each gameplay session is presented in Table 12.

Table 12 *Frequency of corrected errors and phrases identified in the post-play analysis in the* researcher's transcription *of gameplay sessions.*

Category	Error or phrase	Gameplay 1		Gameplay 2	
		Incorrect	Correct	Incorrect	Correct
Corrected errors					
	<person> is telling <the truth>/<a lie>	9	0	5	7
	what's going on?	0	0	0	7
	<person> is suspicious	0	2	0	4
	I saw...	0	15	0	13
	<person> told the truth	0	0	0	1
	what happened?	7	2	0	0
	<person> can't be <job>	0	4	0	0
L1 to L2 translations					
	I know! / Oh, yeah!	0	0	0	0
	There is no...	0	1	0	0
	..., right?	0	0	0	0
Phrases from video analysis					
	I think he is reliable	0	0	0	0
	probably	0	0	0	0
	I have information for you	0	0	0	4
	assume	0	0	0	1

For the **corrected errors** section, it appears that participants were successful in correcting the morphological error related to the verb tell (corrected to "telling"), the correct version of the verb appearing seven times in gameplay 2. Additionally, the phrase "what's going on" which was not used in gameplay 1 appeared frequently in gameplay 2. Related to this is the phrase "what happened" which was uttered incorrectly during gameplay 1 a total of seven times, and not at all during gameplay 2. In addressing the error regarding this phrase, the participants seem to have adopted "what's going on" as an alternative. Subsequently, the use of "I saw" appeared 15 times in gameplay 1, but only 13 times in gameplay 2. It should be noted, however, that gameplay 2 was shorter than gameplay 1 with only 266 total utterances compared to 479 for gameplay 1. Regardless, there was no incorrect usage recorded for the phrase "I saw…" which raises questions regarding why participants initially selected this as a phrase which required further attention. One possible answer is that they highlighted "I saw" as an expression that they deemed important rather than erroneous.

In terms of **translated L1 utterances**, results show that participants did not use any in gameplay 2. However, they did not use the Japanese analogues to these phrases either. One reason for this could be due to their awareness of being assessed as part of the class and were therefore making a conscious effort to not use the L1. In terms of my observations of this group, I can attest to their strong attitude towards boundary formation between L2 and L1 usage for gameplay (L2) and non-gameplay (L1).

Similarly, **words and phrases identified in the video analysis** task appeared infrequently in gameplay 2. Only the expression "I have information for you" being utilized on more than a single occasion. "Probably" was identified as a useful phrase for gameplay 2, however, participants seem to have defaulted to the term "maybe" instead, which was recorded a total of 13 times. In this case then, although participants considered the word "probably" to be something which native speakers use, it was not necessary to progress gameplay.

Discussion

CAF measures: Did oral proficiency improve in the repeated task?

Results suggest that self-transcription and post-task analysis activities helped participants to improve their output accuracy in a repeated gameplay task. This finding matches findings of other, similar studies (Lynch 2001; Stillwell, 2010; Abadikhah & Valipour, 2018). Interestingly, the most frequent errors observed in the initial gameplay task (omission of articles and the grammatical category of number for nouns) were not identified as errors by the participants and thus the same errors appeared frequently in the repeated performance. This emphasises the point both Lynch (2001) and Salas (2015) made: additional teacher support is essential as certain errors may be particularly difficult for participants to identify.

Unlike findings that suggest task repetition promotes more complex performance (Bygate, 1996; Hawkes, 2011; Fukuta, 2015), complexity did not increase in the repeated task in this study even with the addition of post-task form-focused activities. This could imply that the cognitive demands of the task were not reduced due to increased familiarity with the task (see Skehan, 2016). This is to be expected when considering the nature of the task employed in this study: board game play where one's character and thus strategy and winning condition may differ between game sessions. The concept of task repetition in this context then is somewhat troublesome in that although the task is repeated (the same game is played again), participants' in-game identity and goals may differ between games, requiring the use of different strategies and language. Indeed, the complexity of the second gameplay task could have increased for a number of participants regardless of familiarity with the overall task goals.

In terms of fluency, there was no significant difference in mean scores between gameplay 1 and 2. The reason for this could be attributed to multiple factors. First is the above point regarding how roles changed between gameplay sessions. With reference to Table 13, in the first game, R was the godfather and produced the most utterances of all participants (37.78%). R's contribution during the second game was, however, still significant (25.48%). Conversely, in gameplay 2, O was the Godfather and his contribution went up from 1.35% of utterances in gameplay 1 to 15.01% in gameplay 2. This could indicate that the role pushed him to produce more output. Finally, Ku was not the Godfather in either game, yet his contribution was similar for both sessions, indicating his personal disposition to contributing towards the discussion regardless of role. Another unexplored factor in this study is individual participants' perception of task complexity. According to the concept of cognitive load, participants that only produced minimal utterances may have perceived task complexity as too cognitively demanding for them to produce output (Skehan, 2001).

Table 13 *Percentage of participants share of the total number of utterances produced for each game*

Participant	Gameplay 1	Gameplay 2
A	0.70%	4.16%
El	6.36%	8.70%
En	8.06%	3.54%
Ju	6.93%	2.54%
Ko	9.24%	10.08%
Ku	18.91%	23.63%
O	1.35%	15.01%*
R	37.78%*	25.48%
S	10.68%	6.85%

Note: The Godfather role is indicated with a *

In summary then, the fluency mean score did not change significantly between gameplay sessions, however, individual participants' fluency measures may have been influenced by 1) differing game roles, 2) task complexity and 3) participants' motivational characteristics and specifically disposition

to join in discussions. Post task interviews would have illuminated the interplay of these factors more rigorously.

Pedagogical implications

The major pedagogical implications of the current study are that self-transcription was a manageable and effective part of classroom activities (Lynch, 2001) which appeared to support learning gains in terms of output accuracy. However, teacher mediation was required to help learners notice their errors (Mennim, 2007; Salas, 2015). As noted by Foster and Skehan (2013) the activity of transcribing one's performance and listening back to it as a group activity was enjoyable and students were found laughing (pleasantly) at each other's performance and the progression of the game they played. The activity could thus lead to improved group dynamics, and motivation to learn.

Teacher interventions for creating transcriptions

In order to maximize the efficiency and learning potential of the transcription activity, the following interventions are recommended:

- Some students will have never recorded themselves before. **Create a test recording first** to make sure all students can record audio.
- Encourage students to **divide their gameplay audio** between students in order to maximize the volume of data available for analysis. If students do not do this, they can all transcribe the same portion of gameplay meaning that only the first part of their play session is made available.
- **Instruct students on common errors** that occur when transcribing so that their transcriptions are as accurate as possible. This could be as simple as showing the common errors seen in the "Features of student transcriptions" section above: use of the L1, use of ??? to represent an unknown speaker, and transcription inaccuracies (spelling or writing an incorrect word).
- **Instruct students to transcribe verbatim.** Although there were few instances in the students transcriptions here, I have experience of students transcribing *only* the target language that they spoke, leaving any of the L1 out or using automatic translate tools on the L1. Therefore **remind students that transcriptions of the L1 is a useful resource for analysis in the post-task** analysis activity.

Teacher interventions for the analysis sessions

- If students are playing a game that has been assigned and used before in class, the teacher should have data on what errors occur frequently. **Consider creating a worksheet** to guide students on grammar, vocabulary, and other linguistic items to be cognizant of when they are analyzing their errors.
- Encourage students to ask questions about errors.
- Be active in **prompting students to focus on errors** that they have missed.
- If there are many student groups and you are unable to give detailed feedback due to time constraints, **teach students how to search online for the answer to their questions.** Learning about the target language involves more than just looking for one-to-one translations (such as with Google Translate) but reading articles in the L1 on target language usage (i.e. blog posts or language learning communities).

Teacher interventions during the video viewing sessions

Watching videos of target language speakers after students have played meant that students had contextual knowledge before watching, but it is a cognitively difficult task. The following pedagogical interventions are thus recommended:

- **Instruct students to watch through once to understand how gameplay unfolds, then ask them to watch again with a focus on details.** Students need to be encouraged to pay attention to the language used during the gameplay rather than the progression of gameplay.
- **Encourage groups to work together**. Instead of an individual activity, watching **as a group** allowed students to pool their cognitive resources together, and figure out what the target language speakers were saying. In other words, based on my informal observations, if asked to do this task

individually, the cognitive load can be too high and students either burn out and do not watch the video, or they do not rewind and relisten to understand what is being said.
- **Encourage students to ask the teacher for clarifications on what is being said in the video.** Students will find funny, interesting, or key phrases in target language speaker gameplay that they will want to use in their own session (for example: the "I have information for you" phrase in this paper). However, and related to the point above, sometimes they will not be able to parse exactly what the target language speakers are saying. In such cases, the teacher should be available to help.

Limitations of the study

The current study featured a number of limitations which are presented in a bulleted list below with additional explanation provided where necessary.

- Lack of a control group

It is worth considering what would have been found if a control group was employed in this study. That is, would task repetition alone be enough to promote learners to improve their output accuracy, or did the transcription analysis and video analysis tasks play a significant role in improving output performance. Reference to the literature is unclear on this point (see Abadikhah, & Valipour, 2018). Based on Bygate (1996), it could be argued that task repetition alone would promote improved output complexity, no improvement to complexity was recorded in the present study, even with addition of self-transcription.

- Task design

Task repetition in the current study featured the same goals for the group, but due to the nature of the game, participants in-game roles changed between gameplay sessions, and, depending on the role, the amount of speech produced varied considerably. Thus, repetition of this study which defined participant roles to be the same for both gameplay sessions may produce more reliable results.

- Small sample size (n=9)
- CAF measures were limited

As identified by Jackson and Suethanapornkul (2013) there are many ways to measure CAF dimensions. Alternative measures could produce different results, particularly in regards to the assessment of fluency where vocal fluency measures could be adopted.

- Interviews were not conducted post-task

Post-experiment interviews could help illuminate answers to questions that were identified in the study such as why participants decided to add the expression "I saw" to their list of errors.

- Participants' language ability not established before the study
- No measure of subjective task complexity for each participant

The above two points are related. Based on individual English ability, some participants could have been overwhelmed by the task's complexity and therefore unable to output the L2. Additionally, knowing participants' proficiency and task complexity perceptions could also explain why only a few participants made up the bulk of total utterances during gameplay.

Future research directions

Based on the limitations identified above, future research in this area could include more rigorous teacher mediation into the procedure, an element that has been linked to further accuracy improvements (Abadikhah & Valipour, 2018). Additionally, inclusion of a control group (one which does not complete self-transcription and analysis) would illuminate whether repeated gameplay sessions alone are enough to promote improved performance.

Declaration of conflicting interests

I received no financial support for the research, authorship and/or publication of this article.

Acknowledgements

I'm extremely grateful for the reviewers' support (Fred and Vita) in helping me polish this paper into what it finally became.

Massive shout out to the ◺◺◱ crew.

References

Abadikhah, S., & Valipour, M. (2018). The Effect of Self-Transcription and Expert Scaffolding on the Accuracy of Oral Production of EFL Learners. *Pertanika Journal of Social Sciences & Humanities*, *26*(1), 149-166.

Allen, L. K., Crossley, S. A., Snow, E. L., & McNamara, D. S. (2014). L2 writing practice: Game enjoyment as a key to engagement. *Language Learning & Technology, 18*(2), 124-150.

Amiryousefi, M. (2017). The differential effects of collaborative vs. individual prewriting planning on computer-mediated L2 writing: transferability of task-based linguistic skills in focus. *Computer Assisted Language Learning*, 1-21. https://doi.org/10.1080/09588221.2017.1360361

Anderson, T., Reynolds, B. L., Yeh, X. P., & Huang, G. Z. (2008). Video Games in the English as a Foreign Language Classroom. In 2008 Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning (pp. 188-192). IEEE. http://doi.org/10.1109/DIGITEL.2008.39

Bolliger, D. U., Mills, D., White, J., & Kohyama, M. (2015). Japanese students' perceptions of digital game use for English-language learning in higher education. *Journal of Educational Computing Research, 53*(3), 384-408. http://doi.org/10.1177/0735633115600806

Bui, G., Ahmadian, M. J., & Hunter, A.-M. (2019). Spacing effects on repeated L2 task performance. *System, 81*, 1–13. https://doi.org/10.1016/j.system.2018.12.006

Bygate, M., (1996). Effects of task repetition: appraising the developing language of learners' in D.Willis and J.Willis (Eds.) *Challenge and change in language teaching*. London: Heinemann, pp. 136-146.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing*. Harlow: Longman, pp. 23- 48.

Cowie, N. (2018). Student transcription for reflective language learning. *ELT Journal, 72*(4), 435-444. https://doi.org/10.1093/elt/ccy010

Date, M. (2013). The Effect of Task Repetition With Noticing on Proceduralization of Linguistic Knowledge. ARELE: Annual Review of English Language Education in Japan. The Japan Society of English Language Education. https://doi.org/10.20581/arele.24.0_109

deHaan, J. (2019). Teaching language and literacy with games: What? How? Why? *Ludic Language Pedagogy, 1*, 1-57.

deHaan, J. (2020). Game-based language teaching is vaporware (Part 1 of 2): Examination of research reports. *Ludic Language Pedagogy*, 2, 115-139.

deHaan, J., Johnson, N. H., Yoshimura, N., & Kondo, T. (2012). Wiki and digital video use in strategic interaction-based experiential EFL learning. *CALICO Journal, 29*(2), 249-268.

des Pallières, P. & Lamy, L. (2015). *Mafia de Cuba*. Saint-Père-en-Retz, France: Lui-même. Board game.

Enayat, M. J., & Haghighatpasand, M. (2017). Exploiting adventure video games for second language vocabulary recall: A mixed-methods study. Innovation in Language Learning and Teaching, 13(1), 61-75. https://doi.org/10.1080/17501229.2017.1359276

Foster, P., & Skehan, P. (2013). Anticipating a post-task activity: The effects on accuracy, complexity, and fluency of second language performance. *Canadian Modern Language Review, 69*(3), 249-273.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), pp. 354-375.

Franciosi, S. J. (2011). A comparison of computer game and language-learning task design using flow theory. *Computer Assisted Language Learning Electronic Journal, 12*(1), 11-25.

Fukuta, J. (2015). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research*, *20*(3), 321-340. https://doi.org/10.1177/1362168815570142

Hawkes, M. L. (2011). Using task repetition to direct learner attention and focus on form. *ELT Journal*, 66(3), 327-336. https://doi.org/10.1093/elt/ccr059

Hitosugi, C. I., Schmidt, M., & Hayashi, K. (2014). Digital Game-based Learning (DGBL) in the L2 Classroom: The Impact of the UN's Off-the-Shelf Videogame, Food Force, on Learner Affect and Vocabulary Retention. *CALICO Journal*, 31(1), 19-39. https://doi.org/10.11139/cj.31.1.19-39

Hung, H. T., Yang, J. C., Hwang, G. J., Chu, H. C., & Wang, C. C. (2018). A scoping review of research on digital game-based language learning. *Computers & Education*, 126, 89-104. https://doi.org/10.1016/j.compedu.2018.07.001

Jackson, D. O., & Suethanapornkul, S. (2013). The Cognition Hypothesis: A Synthesis and Meta-Analysis of Research on Second Language Task Complexity. *Language Learning, 63*(2), pp. 330-36

Johnson, N., Nelson, M., Lyddon, P., & Selman, A. (2015). Reimagining contemporary EFL curricula. In *JALT Conference Proceedings* (pp. 102-118). JALT.

Jones, D. M. (2019). *Teacher Cognition Related to Tabletop Game Use in Language Learning Classrooms.* (Unpublished masters dissertation). Trinity Western University, British Columbia, Canada.

Koizumi, R. (2012). Relationships between text length and lexical diversity measures: Can we use short texts of less than 100 tokens? *Vocabulary Learning and Instruction, 1*(1), pp. 60-69. http://doi.org/10.7820/vli.v01.1.koizumi

Kormos, J., & Trebits, A. (2012). The Role of Task Complexity, Modality, and Aptitude in Narrative Task Performance. *Language Learning*, 62(2), pp. 439-472. https://doi.org/10.1111/j.1467-9922.2012.00695.x

Levy, M. (1997). *Computer-Assisted Language Learning: Context and Conceptualization.* Oxford: Clarendon Press.

Lintunen, P., & Mäkilä, M. (2014). Measuring Syntactic Complexity in Spoken and Written Learner Language: Comparing the Incomparable? *Research in Language, 12*(4), p. 377-399. https://doi.org/10.1515/rela-2015-0005

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of second language acquisitio*n (pp. 413-468). San Diego: Academic Press

Long, M. (2015). *Second language acquisition and task-based language teaching*. Malden, USA: Wiley-Blackwell.

Lynch, T. (2001). Seeing what they meant: Transcribing as a route to noticing. *ELT Journal*, *55*(2), 124-132.

Lynch, T. (2007). Learning from the transcripts of an oral communication task. *ELT Journal*, *61*(4), 311-320. Retrieved from https://doi.org/10.1093/elt/ccm050

Mennim, P. (2012). Learner negotiation of L2 form in transcription exercises. *ELT Journal*, *66*(1), 52-61.

Miller, M., & Hegelheimer, V. (2006). The SIMs meet ESL Incorporating authentic computer simulation games into the language classroom. *Interactive Technology and Smart Education, 3*(4), pp. 311-328. http://doi.org/10.1108/17415650680000070

Newgarden, K., Zheng, D. (2016). Recurrent languaging activities in World of Warcraft: Skilled linguistic action meets the Common European Framework of Reference. *ReCALL, 28*(3), 274-304. https://dx.doi.org/10.1017/s0958344016000112

Peterson, M. (2012). Learner interaction in a massively multiplayer online role playing game (MMORPG): A sociocultural discourse analysis. *ReCALL, 24*(03), pp. 361-380. https://doi.org/10.1017/S0958344012000195

Peterson, M. (2016). The use of massively multiplayer online role-playing games in CALL: an analysis of research. *Computer Assisted Language Learning, 29*(7), pp. 1181-1194. http://doi.org/10.1080/09588221.2016.1197949

Poole, F. J., & Clarke-Midura, J. (2020). A Systematic Review of Digital Games in Second Language Learning Studies. International Journal of Game-Based Learning (IJGBL), 10(3), 1-15.

Ranalli, J. (2008). Learning English with The Sims: exploiting authentic computer simulation games for L2 learning. *Computer Assisted Language Learning, 21*(5), 441-455. http://doi.org/10.1080/

Reinhardt, J. (2019). *Gameful Second and Foreign Language Teaching and Learnin*g. Springer International Publishing. https://doi.org/10.1007/978-3-030-04729-0

Reinhardt, J., & Thorne, S. L. (2016). Metaphors for Digital Games and Language Learning. In L. Murray & F. Farr (eds.), *Routledge Handbook of Language Learning and Technology* (pp. 415-430). London: Routledge.

Salas, M. D. (2015). Developing the metacognitive skill of noticing the gap through self-transcribing: The case of students enrolled in an ELT training program in Chile. *Colombian Applied Linguistics Journal*, *17*(2), 260-275. https://doi.org/10.14483/udistrital.jour.calj.2015.2.a06

Sample, E., & Michel, M. (2014). An Exploratory Study into Trade-off Effects of Complexity, Accuracy, and Fluency on Young Learners' Oral Task Repetition. *TESL Canada Journal*, *31*(8), pp. 23-46

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, *13*, 206-226.

Shintaku, K. (2016). The interplay of game design and pedagogical mediation in game-mediated Japanese learning. *International Journal of Computer-Assisted Language Learning and Teaching*, *6*(4), 36-55

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press

Skehan, P. (2001). Tasks and language performance assessment. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing* (pp. 167-185). Harlow: Pearson Education.

Skehan, P. (2016). Tasks versus conditions: Two perspectives on task research and their implications for pedagogy. *Annual Review of Applied Linguistics, 36*, pp. 34-49.

Skehan, P. and Foster, P. (1997). The influence of planning and post-task activities on accuracy and complexity in task based learning. *Language Teaching Research, 1*(3), 185-211.

So, M. (2015). *Transcription as an initiator of noticing language form and use: an EFL longitudinal study* (Doctoral dissertation). The University of Toronto.

Stillwell, C., Curabba, B., Alexander, K., Kidd, A., Kim, E., Stone, P., & Wyle, C. (2009). Students transcribing tasks: noticing fluency, accuracy, and complexity. *ELT Journal*, *64*(4), 445-455.

Sykes, J. M., & Reinhardt, J. (2013). *Language at play: Digital games in second and foreign language teaching and learning*. Boston, MA: Pearson.

Sykes, J. M. (2014). TBLT and synthetic immersive environments: What can in-game task restarts tell us about design and implementation? In M. González-Lloret & L. Ortega (Eds.), *Technology-mediated TBLT: Researching technology and tasks*. Amsterdam, The Netherlands: John Benjamins, pp. 149-182.

Wang, Z. (2014). On-line time pressure manipulations. Processing perspectives on task performance. In P. Skehan (Ed.). (2014). *Processing perspectives on task performance (Vol. 5)* (pp. 27-62). John Benjamins Publishing Company.

Yin, R. K. (2018). *Case study research and applications: design and methods.* Los Angeles: Sage.

York, J. (2019a). *English at Play*. Tokyo Denki University Press, Tokyo, Japan.

York, J. (2019b). *Language learning in complex virtual worlds: Effects of modality and task complexity on oral performance between virtual world and face-to-face tasks* (Doctoral dissertation). University of Leicester, UK. Retrieved from https://leicester.figshare.com/account/articles/10242452.

York, J. (2019c). Kotoba Rollers Walkthrough: Board games, TBLT and player progression in a university EFL classroom. *Ludic Language Pedagogy, 1*, 58-114.

York, J., & deHaan, J. W. (2017). Board games and foreign language learning: Rationale and framework development. *The PANSIG Journal 2016*, 379-390. Retrieved from http://pansig.org/publications/2016/2016_PanSIG_Journal.pdf

York, J., & deHaan, J. W. (2018). A Constructivist Approach to Game-Based Language Learning. *International Journal of Game-Based Learning*, *8*(1), 19-40. https://doi.org/10.4018/IJGBL.2018010102

York, J., deHaan, J., & Hourdequin, P. (2019) It's Your Turn: EFL Teaching and Learning with Tabletop Games. In H. Reinders, S. Ryan, & S. Nakamura, (Eds.), *Innovation in Language Teaching and Learning: The Case of Japan* (pp. 117-139). Springer International Publishing. https://doi.org/10.1007/978-3-030-12567-7

Yuan, F., & Ellis, R. (2003). The Effects of Pre-Task Planning and On-Line Planning on Fluency, Complexity and Accuracy in L2 Monologic Oral Production. *Applied Linguistics*, *24*(1), pp. 1-27. https://doi.org/10.1093/applin/24.1.1

Zhou, Y. (2016). Digital vocabulary competition as motivator for learning in CFL classrooms. *Journal of Technology and Chinese Language Teaching, 7*(2), 1-22.

Appendix 1: An example of reduced scaffolding in later task cycles

Part 2: Questions about the rules

In groups, please discuss answers to the following 10 questions about this game

Example:

1. What are the two team colours?

   _____

2. What is 大統領 in English?

   _____

3. The *President* is a member of which team?

   _____

4. What colour is the *Bomber*?

   _____

5. How many rounds are played in this game?

   _____

6. How does the red team win?

   If_____

7. How does the blue team win?

   If_____

8. What can the leader do?

   _____

9. How can I change the leader in my room?

   _____

10. What can I do with my card? (Write three things please)

    _____

An example of teacher mediation and support from the workbook (first cycle of the framework).

Part 2: Questions about the rules

Let's check the rule again. Please write 3 questions about the game rules and then ask your questions to other group members. **Please add questions that your friends ask you.**

Example: *How many minutes is each round? || What is Robber in Japanese?*

Question	Answer
1.	
2.	
3.	
Other group member's question	
Other group member's question	

An example of student-driven exploration of the same activity from the workbook (third cycle of the framework).

Appendix 2: An example of the post task noticing tasks from the workbook

Part 4.1: Correct your mistakes

Look at what players said during gameplay and complete the following sections:

What mistakes were common? What grammar is needed to correct the mistakes?

Grammar point	Example
<person> should <verb> <object>	*You should collect the red cards*

What Japanese expressions did you use a lot? How can you say it in English?

Japanese	English
どうしようかな	*What shall I do?*

Watch YouTube examples of natives playing. What expressions do they use? Please work with your group to understand what they are saying.

English	Japanese translation