Article:
Bui, G, Ahmadian, MJ orcid.org/0000-0003-3458-1854 and Hunter, A-M (2019) Spacing effects on repeated L2 task performance. System, 81. pp. 1-13. ISSN 0346-251X

https://doi.org/10.1016/j.system.2018.12.006

Spacing effects on repeated L2 task performance

Gavin Bui, The Hang Seng University of Hong Kong
Mohammad Javad Ahmadian, University of Leeds, UK
Ann-Marie Hunter, University of Leeds, UK

Abstract

This study aimed to explore task repetition (TR) under five performance conditions with different intervals between the initial and repeated performance. 71 adult learners of English as a foreign language performed a picture description task and, according to their randomly assigned groups, repeated the same, unanticipated task with either no interval (immediate repetition), a one-day, a three-day, a one-week, or a two-week interval. Performance was assessed using a range of measures which capture the three dimensions of complexity, accuracy, and fluency (CAF). Results showed that, overall, TR had a positive effect on L2 performance regardless of the length of intervals. Spacing appeared to mediate the effects of TR in terms of fluency and structural complexity with speed fluency benefitting most from immediate or small intervals between initial and repeated performances. Structural complexity and repair fluency scores were higher with an interval of one week between performances. Findings are discussed in terms of underlying speech production processing and the implementation of TR in the language classroom.

1. Introduction

A growing body of research has demonstrated that task repetition (TR) could positively affect both L2 oral and written task performances (e.g. Ahmadian & Tavakoli, 2011; Amiryousefi, 2016; Bygate, 2001; Lambert, Kormos & Minn, 2017; Thai & Boers, 2016; Wang, 2014). TR is widely used in L2 classrooms and, contrary to the lay view, language learners do not find repeating the same or slightly altered tasks redundant or mundane but believe that it is quite beneficial for improving their L2 performance (Ahmadian, et al. 2017; Lambert, et al. 2017). Research on the effect of TR has mainly focused on whether and how it affects L2 oral performance in terms of complexity, accuracy, fluency, and lexis. A review of the available literature reveals that the studies conducted so far have had different foci: while some have investigated the extent to which TR impacts on performance *of a new task*, that is asking participants to repeat a task and then requiring them to do a new one (e.g. Kim & Tracy-Ventura, 2013; van de Guchte et al, 2016; de Jong & Perfetti, 2011), the vast majority have looked at the impact of TR on performance of the *same* task (e.g. Bygate, 1996; 2001; Wang, 2014; Boers, 2014; Thai & Boers, 2016; Lambert et al., 2017; Gass et al, 1999; Lynch & Maclean, 2000; 2001). The findings of this latter body of research are broadly consistent, demonstrating gains for fluency and more variable findings for complexity and accuracy on the repeated task performance.

In most of the above-mentioned studies, the theoretical framework employed to explain these gains has been Levelt's (1989) speech production model. This model, which was initially proposed for L1 speech production, has proved particularly useful for explaining the ways in which manipulating task performance variables affects the

underlying speech production processes (de Bot, 1992; Kormos, 2006). The three main information processing stages posited in the model include Conceptualization, Formulation and Articulation. Conceptualization is the stage in which the speaker decides on and conceives the message which is to be communicated. The outcome of this stage is conceptual and non-linguistic which Levelt labels 'pre-verbal message'. This non-linguistic message will then feed into the next stage, i.e. Formulation, in which the appropriate lexical and syntactic elements are selected and are mapped onto the preverbal message to produce what is called 'internal speech' or 'phonetic plan'. In the third processing stage, Articulation, the phonetic plan is converted into the overt speech (de Bot, 1992, p. 1). Conceptualization is an entirely conscious and cognitively demanding process for both L1 and L2 speakers – we all think about what we want to say before we say it. But, whereas formulation and articulation stages are fairly automatic and effortless for L1 speakers (Levelt, 1989; Kormos, 2006), they could prove particularly demanding for L2 speakers. This is because, accessing and retrieving the relevant lexical and syntactic information, which are to be mapped onto the intended message, exert much more cognitive demand on L2 speakers than native speakers of a language. As such, L2 speakers have to divide their limited attentional resources among the three stages of speech production (Ahmadian & Tavakoli, 2011; Skehan, 2014).

But, how does this allocation of attention take place in the context of task repetition? Previous studies cited above have confirmed that since performing a communicative task requires L2 speakers to process meaning first, during their initial encounter with a task, speakers tend to prioritize the conceptualization stage (i.e. determining what to say) over formulation and articulation. However, during a repeated encounter with the same task, they will be able to, somehow, skip the conceptualization stage because they already know what they want to say and allocate their attentional and monitoring resources to formulation and articulation. This could result in more fluent, accurate, and complex language. It should be pointed out in passing that complexity, accuracy, and fluency have been argued to reflect fluctuations in learners' attention to language during communication and that a balance of foci is commonly assumed to be related (indirectly) to balanced instructed SLA (see Skehan, 1998). One important, yet neglected, aspect of the operationalization of TR is the spacing between the original and repeated performances. This has to do with whether a task is repeated immediately or after a day, a week or a month. We will now turn to this topic.

2. Massed vs. distributed task performance

More than a hundred years of research into distribution of practice in educational psychology suggests that variations in spacing and distribution of practice does make a difference (Carpenter, et al. 2012). In addition, this body of research suggests that, overall, "knowledge is retained better when the practice is distributed rather than massed" (Suzuki & DeKeyser, 2017, p. 2). This phenomenon is generally known as "spacing effect" which refers to "to enhanced memory performance on repeated items whose presentations are distributed (either through time or through time and other presentations), as opposed to performance on items whose presentations are massed or contiguous." (Glenberg, 1979, p. 95). There is a wealth of

research evidence in applied cognitive psychology supporting the benefits of distributed practice for learning and retention of verbal information (e.g. Bloom & Shunell, 1981; Cepeda, et al. 2006). There are two main theoretical explanations for the effects of distributed practice. According to Encoding Variability theory (Glenberg, 1979), distributed practice results in better recall because on each occasion of practice participants encode the taught materials differentially and, therefore, will have various retrieval cues (Serrano, 2011). Deficient Processing theories claim that some 'lag' or interval is required for the presented items to be sufficiently processed in order for further retrieval to take place efficiently (Serrano, 2011).

Such research is scarce in TBLT. In the SLA literature, too, the results are mixed and inconclusive. Most of the studies into the effects of intensive versus extensive language teaching programs have either found no statistically significant difference or have provided evidence in support of the superior effects of intensive instruction. For example, Serrano (2011) who investigated 152 students (18 to 23 years old) found no statistically significant difference between intensive and extensive instruction in terms of proficiency development (also see Serrano & Muñoz, 2007). In a recent study, Suzuki & DeKeyser (2017) looked at 40 beginning level learners of Japanese as a second language and found that, compared to distributed practice, massed practice could result in more fluent (rapid) and accurate production of morphological features. In this study, distribution of practice was operationalized along a continuum of 1-day, 7-day, and 21-day intervals. The apparent disparity between results from instructed SLA and cognitive psychology could be mainly attributed to the differences between the nature of the target features to be learnt and the design of the experiments (Serrano, 2011; Suzuki & DeKeyser, 2017).

Although, to the best of our knowledge, no study has empirically investigated the impact of spacing in TR, it may still be informative to compare the results of studies that have used different repetition intervals. Some TR studies have looked at what happens when the task is repeated *immediately* (Boers, 2014; Lambert et al., 2017; Lynch & Maclean, 2000; 2001; N. de Jong & Perfetti, 2011; Thai & Boers, 2016), while others have investigated repetition after an interval of a few days (e.g. Bygate, 1996; Pinter, 2005), weeks (e.g. Fukuta, 2016; Hsu, 2017) or months (Bygate, 2001; Azkarai & Garcia Mayo, 2016). Table 1 shows that, in general, fluency seems to increase as a result of TR *regardless* of the interval between the original and subsequent performance(s). This suggests that fluency is fairly robustly affected by TR. The picture with accuracy and complexity, on the other hand, seem to be less clear. In general, it appears that a short interval between original and subsequent performance makes it more likely that an effect will be seen for accuracy and complexity. A possible explanation is that if the repetition intervals are very short (i.e., immediately after the first occasion of task performance), speakers will find it easier to draw on recently activated and retrieved lemmas which will increase the speed of processing. This hypothesis is in line with de Jong's (2012) argument that "immediate repetition makes it more likely that benefits of conceptualization and formulation persist into the repeated deliveries" (de Jong, 2012, p. 44). Therefore, it seems plausible to suggest that if certain speech production processes are speeded up by recency of use, then there may be more attentional resources

available for monitoring processes. This could in turn promote complexity and accuracy. However, accuracy and complexity of speech might not be affected to the same degree as fluency (Skehan, 1998, 2009, 2014, 2018). Previous research has demonstrated that in order to observe significant increases in terms of complexity and accuracy from the first to the second occasion of task performance, learners need some time to fully integrate the new form-meaning pairings into their long-term memory and then compare them (consciously or unconsciously) against the target-like forms that they might have in their declarative memory or with those that they may come across in the input (see, Housen, et al. 2012). Thus, one could speculate that if the time interval between the two occasions of task performance is relatively longer (say, by a few days), learners might be less fluent/more hesitant but more accurate/complex precisely because they have had ample time to process the content of the task and check the accuracy of form-function mappings consciously. This will in turn induce them to self-monitor and self-correct their speech which could lead them being more accurate/complex but less fluency (Kormos, 2006). These speculations could be supported by the findings of Lambert, et al. (2017). This might also explain why effects for accuracy and complexity are often only seen on the third performance (e.g. Boers 2014; Thai & Boers, 2016).

An outstanding issue is that the studies that have examined immediate repetition have tended to look at multiple repetitions and therefore the finding that all areas of performance are affected may be related to *number* and not *spacing* of performances. Clearly, then, further research is needed to investigate whether spacing alone mediates the impact of TR and whether all aspects of performance are affected similarly.

3. Research questions

This study aims to explore the effects of task repetition on task performance, and how such effects, if any, are subject to the influence of five spacing conditions between the initial and the repeated task. Two research questions guided this research:

RQ1: What are the effects of task repetition on L2 oral performance in terms of complexity, accuracy, and fluency?
RQ2: In what way does the spacing of repeated performance mediate the effects of task repetition?

4. Methodology

4.1. Participants

71 undergraduate students participated in this study. They were selected from a larger pool of students in a non-credit bearing English enhancement course at a private college in Hong Kong. The students who participated in this study were first-year or second-year students, aged between 18-21 (mean=18.87, *SD* = .97). They were all language majors and had results in the English Language Examination (for College Entrance in Hong Kong) ranging between 3 and 4, which are roughly equivalent to IELTS bands 5.5-6.0. Finally,

none of them had stayed in English-speaking countries for more than three months. The participants were randomly assigned to five different task repetition groups (14-15 students per TR group, see Table 2). To further ensure that the proficiency levels were comparable across the groups, a proficiency test consisting of four reading comprehension articles and four listening passages (40 questions in total) from an IELTS test mock paper were administered before the task. The ANOVA result showed that there was no proficiency discrepancy between any two groups (F = .226, p = .923).

Table 2 about here

4.2. Design

This study used a two-way design. The first variable was 'repetition' with two levels (repeated measures) – first and second performance (n = 71). The second variable was spacing with five levels (between subject) (n = 14/15).

4.3. Materials

A picture description task (see the Appendix) was used in this study. The rationale for using this single-frame picture was to reduce the interference of irrelevant factors, such as pre-planning time (a picture series would inevitably require much more time for viewing and comprehending) and cognitive skills (e.g., ability to make logical connections), which were not the amongst the variables of this study. The current picture task allows participants to start speaking as soon as they see it as it involves a simple and familiar story at a household setting. On the other hand, there is a certain degree of suspense and intensity (the lightbulb is falling while the man is trying to catch it), which pushes the participants to elaborate and present their versions of the story.

4.4. Procedures

All participants were instructed to perform the picture description task described above. This was the third task in the lesson. The other two tasks served as warm-up activities so that students could get used to the recording device and feel more relaxed for the target task. In this task series, the first task was a pronunciation training task in which the teacher reviewed the consonants and vowels taught in the last lesson, and introduced five new sounds, followed by a picture naming task which involved the target sounds. The second task was a 3-minutes cartoon description task, followed by a performance in front of the class by 3 to 4 students randomly chosen as a post-task activity. Then came the third one, the picture description task, which preceded yet another public performance by 3-4 students (other than those who had performed in the second task). An important note here is that, to encourage students to do their best, the students were reminded that their recorded performances would be used as part of the course assessment. However, they were not informed that they were supposed to do the task again until the second iteration. The second performance came as a surprise. This was to distinguish task repetition as task-internal readiness from dry rehearsal which is a kind of task-external readiness (Bui, 2014).

14 participants immediately repeated the same task after the first one. The one-day group (n = 14) performed the same task after a one-day interval (The first and the second tasks happened on the first and the third day respectively). The third group (n = 14) did the same but the interval between tasks was three days. The fourth group (n =15) repeated the same task with a one-week interval. The last group (n = 14) had a two-

week break before repeating the same task. The spacing conditions from no interval to two weeks were chosen to explore spacing which was also pedagogically meaningful. A two-week limit was set because it was felt that any longer interval would be untenable for language teaching purposes. Except for different intervals between the initial and the repeated task, all participants followed the same instructions and procedures. Then, all participants in the same TR group had the regular English enhancement lesson taught by one of the authors, which was assisted by a research assistant. They were instructed to record their speaking tasks during the lesson with their own mobile phones. The picture description task was the third speaking task in the lesson so that the participants could get used to recording and the task procedure. After the "intended" task, the lesson went on as usual and the participants were not informed that the same task would appear immediately or after an interval again. For the immediate repetition group, the participants were requested to perform the task again right after the previous task in the same class. Participants of the three-day, one-week and two-week groups performed the same picture description task a second time following the same procedure of their first attempt in their regular lessons. Participant of the one-day interval group had a special "make-up" class because the one-day interval did not fit in their regular lesson schedule, but they also followed the same procedure.

4.5. Analysis

Task performance has been commonly measured in terms of complexity, accuracy, and fluency (CAF, see Bui & Skehan, 2018). One reason why CAF measures have been used in various studies (including the current study) is that the CAF triad has been shown to be amongst the most useful and valid indicators of the facets of L2 performance. Housen, et al. (2012) go so far as to suggest that, with the cognitive turn in SLA literature, "CAF have also started to figure as central foci of investigation in their own right" (p. 2). They also report that, in several instructed SLA studies, the CAF triad emerge "as the primary epiphenomena of the psycholinguistic processes and mechanisms underlying the acquisition, representation, and processing of L2 systems" (Housen, et al. 2012, p. 2). The main justification for employing CAF measures in the current study was to enable us (and, of course, future researchers) to interpret and discuss the results of this study in light of previous studies on task repetition.

In the literature, complexity, accuracy, and fluency are construed as multicomponential constructs (Housen & Kuiken, 2009). Therefore, it is imperative to use complementary but distinct measures which tap into various dimensions of the underlying constructs – this will help to avoid redundancy in measurement (Norris & Ortega, 2009). However, although measures used to assess CAF sub-constructs do not always show differences among groups of participants, they still may prove scientifically valid and informative about the underlying construct of language proficiency. This is related to what Pallotti (2009) labels 'the necessary variation fallacy'. In the meantime, we need to be mindful of the fact that fluctuations in CAF triad (or any other aspect of L2 systems) are dynamic and non-linear. Therefore, an important caveat in interpreting the results of any study which uses these measures is that: improvement in CAF measures should not be taken as development of the L2 system.

This study, therefore, follows this CAF convention and, following previous studies, includes a total of twelve indices as shown in Table 3. For example, both structural and lexical complexity measures were used as they have been shown to measure different aspects of linguistic complexity and in fact complement each other (Skehan, 2009). Accuracy ratio and errors per 100 words were adopted as indices of accuracy as the former serves as the overall accuracy rate based on a production unit (a clause) and the latter is thought to capture all errors that occur (Bui, 2014). Following Tavakoli & Skehan (2005), this study distinguishes between speed, breakdown and repair fluency. The *speed* fluency measures selected here is speech rate. Although speech rate is often considered a 'global' or 'composite' measure of fluency (Huensch & Tracy-Ventura, 2017), it provides adequate indication of the overall speed with which a person speaks. Following developments in the study of breakdown fluency (Bosker et al, 2013; Tavakoli, 2015), *breakdown* fluency focuses on mid-clause pauses which are more likely to be indicative of L2-specific fluency. In terms of *repair* fluency, we took into account frequency of phenomena such as reformulation and false starts. In most cases we used pruned text; however, some other measures such as repair dysfluency measures require unpruned transcriptions. There is a limitation in relation to measurement of fluency which needs to be acknowledged at this juncture. Although, as an anonymous reviewer of System has rightly pointed out, end-of-clause pauses are of paramount importance in the measurement of fluency, owing to the limited scope of our analyses we did not take them into account. However, the data will be available for future researchers to look into this.

We used the GoldWave software which converts sounds into spectrograms. This allowed us to identify pauses of up to 0.01 second. Following Foster & Skehan (1996) and most other Skehan's studies (see Skehan, 2018), we used 0.4 seconds as the length of noticeable pauses (also see Bui, 2014 and Bui & Huang, 2018 studies). All data was co-coded by the first author and research assistants (RAs), and they reached 100% agreement on the scores after discussion. The RAs made the initial coding of the transcribed task performances, and the first author checked all of the coded files again. In case of disagreement, they had a discussion until they came to a consensus. We thought this would be a more logical and practical attempt at ensuring maximal accuracy in scoring than completing all scoring and then calculating reliability post hoc.

Table 3 about here

5. Results

We used General Linear Model's repeated-measures test on SPSS to analyze the data (see Table 4 for Wilkins' Lambda values for all variables). Although we have mainly drawn on the Null Hypothesis Testing (NHST) model to analyze and interpret our data, it is important to acknowledge that this model of dichotomous thinking is deeply flawed and incompatible with the complexities and intricacies of the phenomena that Instructed SLA researchers are interested in (see Plonsky, 2015 or, better still, Cumming, 2012 for a full account of the limitations of NHST thinking). Therefore, to partly address this problem, we have reported effect sizes for all mean differences (whether statistically significant or not) and have attempted to point out which significant or

insignificant differences are plausible and reliable in light of our sample size and post-hoc power estimates. Also, all confidence intervals will be supplied as supplementary material and the whole dataset will be available for further analyses and examinations. This is aimed to be a step towards transparency and open science and will be helpful for both future empirical studies and metanalyses on this topic. The main issue to bear in mind in interpreting the results of this study is that, due to rather small sample sizes in each group (i.e. the spacing variable), insignificant results should not necessarily be taken to mean that spacing and repetition do not interact with each other. Cumming (2012) clearly demonstrates how a slight change to the sample size and/or to the variability within the sample could lead to (in)significant results. Therefore, following Cumming (2012, p. 9), in what follows we will attempt to be as explicit as possible about the likelihood of Type II error as well as highlight "the uncertainty inherent in our data".
Table 4 about here

Complexity

Table 5 shows that, task repetition *per se* leads to higher structural complexity but the difference between different spacing conditions is not statistically significant. This could very well be due to the size of the sample (n = 14/15) and the fact that a sample of this size does not have sufficient statistical power to allow rejecting the null-hypothesis. In other words, results could have been statistically significant if we had, for example, a few more participants and/or less variability (smaller SD) in our sample (Cumming, 2012). Results also indicate that task repetition pushed learners to use more subordinate clauses (p = .00, η^2=.21) and, more words (p = .00, η^2=.19), within an AS-unit. However, lexical diversity as indexed by *D* did not seem to be influenced by task repetition (p = .26, η^2= .09). The sample size for repeated-measures variable is 71 which affords sufficient power to reject the null-hypothesis. As for the effect of spacing between the initial task and the repeated task, no significant results were found in either structural or lexical complexity, though it was approaching significance in Words per AS Unit (p = .06). Here again the observed power is .65 whilst the ideal observed power is .80. Therefore, it could be suggested that the sample has not afforded sufficient power to reject the null-hypothesis. The biggest gain in length of the AS unit seemed to happen with a one-week interval (mean difference: 1.35). In sum, our analyses revealed no interaction effects between task repetition and spacing and it appears that, regardless of the spacing of performances, repeating the same task raises structural complexity but not lexical diversity.
Table 5 about here

Accuracy

Table 6 reports the results for accuracy of performance. As is shown in the table, repeating the same task did not result in increased ratio of accurate clauses. A closer inspection of the power estimates reveals that the observed power for repetition is only .14 and for the interaction of spacing and repetition is only .19 which do not seem adequate for rejecting the null-hypothesis. Therefore, the likelihood of Type II error should be seriously considered in interpreting this finding. However, as displayed in

Table 6, the number of errors has significantly decreased (p = .01), albeit with a small effect size (η^2=.09). In terms of the spacing between repeated performances, there was no significant effect on accuracy. Neither was there any interaction effects between task repetition and time intervals. But here again, the observed powers were rather small (.29 for the effect of intervals and .27 for the interaction between spacing and repetition). Therefore, the data obtained from our sample seems to show that the sort of task repetition explored in this study has, at most, a subtle influence on accuracy, and that spacing conditions do not moderate the effect.

Table 6 about here

Fluency

As can been seen from the results below (Table 7), both task repetition and spacing significantly affected speech rate. In general, learners had a higher rate of speech in the second performance (p = .01, η^2= .11, observed power = .80), and this increase in speed fluency was mediated by the spacing between the initial and the repeated task (p = .03, η^2 = .15, observed power = .75). To further determine the spacing effects, an LSD *post hoc* test was performed which showed that immediate repetition (the control) significantly outweighed all other intervals (one-day: p = .02, one-week: p = .03, and two-week intervals: p = .01), except for the three-day interval (p = .59). The one-day interval, however, did not prove significantly different from the three-day (p = .07), one-week (p = .86), or two-week intervals (p = .75). Task repetition with a three-day interval was significantly better than that with a two-week interval (p = .04) in terms of speech rate, but not better than repetition with a one-week interval (p = .10). There was no difference in speech rate between the one week and two weeks (p = .62) intervals.

For breakdown fluency, task repetition led to a decrease in the number of pauses in the middle of a clause which was approaching significance (p =.06, η^2 = .05, observed power = .46) but the spacing of repeated tasks did not have any effect (p = .56, η^2 = .04, observed power = .23). In contrast, task repetition did not seem to influence the overall frequency of filled pauses in participants' speech, and the significant differences between the spacing conditions at the outset of the study suggests that filled pauses like "er" and "hmm" might be more related to individual speaking style. In particular, an interval of one week resulted in greatest reduction of filled pauses in comparison to all other TR conditions (p = .00 compared with immediate repetition, one-day, and three-day intervals, p = .02 compared with the two-week interval).

Table 7 about here

Similar to the findings concerning filled pauses, there were very limited effects of task repetition and spacing of performances on repair fluency, as displayed in Table 8. TR did not result in changes in the incidence of repairs when the combined effects of all five spacing conditions were taken into account. An interval of one week, however, seemed to slightly reduce the number of verbatim repetition (p = .05, η^2=.12) but there was no spacing effect on the other three repair measures.

Table 8 about here

6. Discussion

A general pattern that one could identify in Tables 5 to 8 is that when learners repeat a task there is an effect on their performance in terms of CAF irrespective of whether the repeated task comes a minute or a week after the original. This supports the findings of Wang (2014) and others which show similar effects for TR. However, the interesting additional finding is that, in terms of fluency specifically, the *spacing* of the two performances mediates the effect of TR with greater gains at smaller intervals. By contrast, structural complexity seems to be most affected when the task is performed with an interval of one-week. The following discussion will try to interpret these results in relation to Levelt's (1989) speech production model as well as other relevant frameworks such as Skehan's (2014) Limited Attention Capacity hypothesis. However, as it was explicitly stated in the results section, some of our comparisons are not entirely reliable owing to our small sample and insufficient power to reject the null hypotheses. Especially, in relation to spacing effects (where the sample size was 14/15), the likelihood of Type II error seems to be high. This is mainly because there is cogent theoretical rationale (as discussed in the literature review section) to justify positing a mediating role for spacing. Further research is needed to explore this issue.

6.1. Task Repetition

In response to RQ1, the results of this study showed that task repetition promoted structural complexity but not lexical complexity (as indexed by *D*). This finding lends support to some other TR studies (e.g., Bygate, 2001; Lynch & Maclean, 2001; Wang, 2014). What is interesting is that many other TR studies have found that a single repetition did not impact significantly on structural complexity (e.g. Boers, 2014; Thai & Boers, 2016). It would be interesting to investigate why such a contrast occurs. Drawing on Levelt's (1989) model, one plausible explanation for this increase might be that familiarity with the task, and the reduced processing load required to engage in *conceptualization,* frees up resources which can be devoted the formulation, articulation and self-monitoring stages of speech production. This is reflected in a higher ratio of subordinate clauses and also longer sentences (in words) during the second task performance, as learners are more able to integrate their (more advanced) declarative knowledge into real-time communication. This change will then be realized through more elaborate language (lengthier stretches of discourse) and more sophisticated structures (higher amount of subordination). Further (qualitative) analysis of the speech data may prove fruitful in further interpreting this finding. Language learners' ability to integrate more complex language into their performance may benefit their interlanguage re-structuring and development. This issue, of course, remains to be tested empirically. More specifically, future research could look into to the extent to which changes in complexity are retained over time or are transferred to the performance of an entirely different task.

A surprising finding was the lack of influence of task repetition on lexical diversity. Given that TR in the present study seemed to promote structural complexity, it follows that learners might also be more successful in searching for a wide range of lexical items, and as a result produce a more lexically rich and diverse output. However, the current study does not seem to support this hypothesis. We are inclined to attribute this to our participants' rather limited L2 lexicon. As it was stated in the Methodology section,

participants were first and second year non-English majors with lower-intermediate to intermediate L2 proficiency. Learners at this level are normally characterized by a mental lexicon that is smaller, less organized and likely slower in access (Skehan, Foster, & Shum, 2016). Alternatively, and as suggested by Boers (2014), Thai & Boers (2016) and Hunter (2017) it may be that additional repetitions may be necessary to provide learners with the opportunity to take advantage of increased monitoring at performance two and then make qualitative changes to their speech during performance three.

Task repetition had a small but, nonetheless, significant impact on one of the accuracy measures (errors per 100 words) though not on the other measure (i.e., ratio of accurate clauses). As explained above, the results of task repetition studies with regard to accuracy are quite mixed. Studies such as Bygate (2001), Hsu (2017), Sample and Michel (2014) found no accuracy effect for a single task repetition. A growing body of research is now suggesting, however, that three or more iterations of a task may be necessary in order to bring about changes to accuracy (Boers, 2014; Thai & Boers, 2016; Hunter, 2017). Alternatively, it may be that TR combined with some other sort of task-readiness (Bui, 2014; Bui & Teng, 2018) could result in higher accuracy (e.g., Ahmadian & Tavakoli, 2011 with repetition + careful online planning; Hsu, 2017 post-task transcription + repetition; Wang, 2014 with pre-task video watching + repetition condition). It may be, then, as hypothesized by Bygate (2001), Sheppard (2006) and others that learners require a feedback or reflective stage between initial and repeat performances in order to make use of freed-up processing capacity to make improvements to their performance in terms of accuracy.

As for the effects of task repetition on fluency, firstly, speech rate increased significantly between the first and second performance of the task. This is now a very robust and consistent finding in TR research. We take this as support for both planning and priming explanations of TR. TR impacts on oral fluency are that the first performance provides learners with the opportunity to 'plan' the content for the next performance and also that the language produced is 'primed' and can be more speedily activated on a subsequent occasion. Priming is an especially attractive theoretical explanation because of the substantial amount of verbatim repetition that happens during task repetition (Boers, 2014) especially when the repetition is immediate.

In terms of breakdown fluency, the results show that the number of mid-clause pauses were lower during the second performance overall and this was approaching significance (p = .06, η^2=.05), suggesting that task repetition reduces the need of the learner to pause online to conduct formulation. This is in line with Lambert et al (2017) in that a single repetition was insufficient to impact significantly on mid-clause pauses. It appears then, as was suggested in Lambert et al (2017) that additional iterations of the task may be necessary to have significant impact on this aspect of fluency. Filled pauses were impervious to the effects of a single repetition of the task. These areas of fluency may be part of learners' idiosyncratic speech features (or personal styles in Skehan, Foster & Shum's (2016) terms) that seem to be more resistant to task influences.

There was no TR effect observed for the repair measures, which supports Lambert et al's (2017) finding that a single repetition is insufficient to impact on this aspect of fluency. It also has to be mentioned that, except for reformulation, the number of

repairs in this data-set was generally very low, and therefore the statistically (non-)significant results should be treated with caution.

6.2. Spacing

RQ2 is concerned with the mediating effects of spacing on TR. Although spacing between repeated task performances did not have major effects on all measures of performance, the effects of TR on some complexity and fluency measures were mediated by the length of interval. As it was emphasized in the results section, the likelihood of Type II error should not be under-estimated and future research is necessary to see the extent to which our findings are reliable. Table 9 below summarizes these moderating effects. It appears that speech rate increases most if the interval between the two performances is very short (immediate), but structural complexity and some repair measures best improve if there is a one-week interval. As the results of this study show, one's speech rate is most sensitive to the distribution of performance occasions. This effect is lessened when the interval increases. The swifter access to ideas and the linguistic resources in terms of lexis, grammar and phonology eases the articulation of a linguistic plan and greatly alleviates the demand for online planning. This finding is in alignment with the results of Lambert, et al.'s (2017) study and supports the general findings of previous SLA research on distribution of practice which suggests that massed practice tends to be more advantageous for language development (Suzuki & DeKeyser, 2017). When the repetition is immediate, learners can draw on language that is readily 'primed' for usage. Indeed Boers (2014) and Thai & Boers (2016) found that during immediate task repetition, learners were able to memorize and regurgitate exceptionally long chunks of speech *verbatim*. The longer the interval, the lower the chances of lexico-grammatical items to remain in a readily accessible state for learners to retrieve them. With longer intervals, it could be assumed, one has to start all over again and replicate part (if not all) of conceptualization and formulation. This, in turn, means that the overt speech will be noticeably less fluent.

Table 9 about here

By contrast, three measures (words per AS unit, filled pauses and repetitions) were affected most over a one-week interval. These may be related to the nature of the measures and the underpinning speech production processes to which they relate. Spacing of a week may afford learners the opportunity to encode the ideas and concepts adequately (Glenberg, 1979, for his encoding variability theory). After a week, it is less likely that one could retrieve a large amount of verbatim language from the previous performance especially when learners are unaware that they will repeat the task; therefore, learners must re-formulate their message, but the residual effect of the first occasion is still strong enough to be facilitative. With such a residual effect, one is primed and can further build on the previous speech and move from a "what to say" state in the initial task to a "how can this be said better" state in the subsequent performance. This may lead to more elaborate language as evidenced by longer AS units. Similarly, filled pauses and verbatim repetition are often means to "buy time" for lexical retrieval or syntactic assembling. An interval of one week enables learners to sufficiently process the previous task information and reduce the need for frequent repairs and

pauses for lemma retrieval and grammatical encoding (Serrano, 2011). All this may point to the pedagogical implication that spacing task repetition at a one-week interval is a desirable means to improve complexity and repair fluency, whilst immediate repetition tends to benefit speed fluency.

In contrast, accuracy was not found to be related to the spacing between performances. As noted above, it may be that a single repetition was insufficient to observe such changes. Another possibility is related to one's proficiency level, a more inherent characteristic among L2 learners and that these learners were unable to make significant improvement in a few days or weeks.

6.3 Theoretical and pedagogical implications

The task-based literature has generally confirmed the facilitative effects of repeating the same or slightly altered task on L2 performance. What remains speculative in this area is the spacing effects that might accrue from repetition. Previous studies have adopted research designs of one or multiple repetitions with intervals ranging from immediate repetition (e.g., Boers,2014) to ten weeks (e.g., Bygate, 2001). No research has directly compared the effects of different spacing conditions in a single study. This research is probably the first attempt in this regard, and we did find that different distributions of repeated performances along a timescale influenced the areas of breakdown fluency, repair fluency and structural complexity. Yet, this could very well be due to the small sample size and the rather high probability of Type II error. As it was shown in the results section, the observed power for non-significant results were quite small. Despite all this, three theoretical implications could be drawn from our findings. First of all, massed or distributed task performances appear to exert different impact on performance. Early (e.g., Bloom & Shunell, 1981) or more recent (Cepeda et al., 2006) applied psychological research has, in general, lent support to the beneficial effects of spaced rather than massed practice for the learning of verbal information. In contrast, this study further suggests that massed performances might be more beneficial for L2 performance, at least in terms of fluency and complexity, than distributed practice. Secondly, the effects of TR, in and of itself, is robust, and may sometimes negate the possible effects of spacing. This echoes many past studies, such as Bygate (2001) which had a span of ten weeks between the two performances, that repeating the same task contributes to the improvement in various areas of L2 performance. The third issue concerns task assessment methodology in complexity and accuracy. It has been shown in this study that structural and lexical complexity measures distinctive aspects of task performance. They should be employed jointly in a study and neither of them should be viewed as synonymous to the notion of complexity. Similarly, there are situations where some measures are more sensitive than others in capturing accuracy of task performance. For example, though both being indices of overall accuracy, the "number of errors per 100 words" measure in this research seems to work better in detecting the change of error counts between repeated performances than the more commonly used "ratio of error-free clauses". Again, the differences in measurement may be in part responsible for inconsistent findings reported in the TR literature.

Results of this research may have some pedagogical implications. The immediate impression we gain from the findings is that TR makes quite a useful and powerful

means of increasing CAF in L2 speaking performance. Furthermore, there is growing evidence that learners do not find repeating the same task boring or redundant (as evidenced in Ahmadian et al., 2017; Lambert et al, 2017; Lynch & Maclean, 2000); rather, they feel it is facilitative. This all adds to the argument for the inclusion of TR in a language teachers' arsenal of classroom activities. Another interesting implication pertains to how repeated tasks should be distributed. If a teacher wishes to foster increased fluency in practice, repetition with shorter intervals might be more conducive. We would also suggest that this increased fluency may reflect fluency of underlying cognitive processing and might mean that learners, given the opportunity to further repeat the task would continue to make changes to their performance (in terms of accuracy and complexity). Immediate repetition is reasonably straightforward to implement in the language classroom through the use of 'inherently repetitive' activities such as Maurice's (1983) 4-3-2 technique and Lynch and Maclean's (1994) 'Poster Carousel'. This study has also suggested that there is a case for repetition which takes place over a period of one week. This spacing of repetition might facilitate variation and structural change in performance (hence structural complexity). Overall, a combination of tasks which include some immediate repetition and those with a weekly interval can help contribute to the improvement of L2 performance.

7. Conclusion

Task repetition as a pedagogical means to improve task performance has been extensively studied but its effects relative to spacing between repeated performances seem less touched upon in the literature. This study aimed to fill this gap by examining how different spacing conditions may have an impact on task repetition effects. It was found that, regardless of the length of spacing, learners spoke with more complex utterances and with greater speed and fewer incidences of breakdown during a repeated performance of a task. They were also slightly more accurate during the second performance but showed no improvement on lexical diversity or repair fluency measures. In contrast, the spacing between performances seemed to have some impact on structural complexity and fluency. In general, an interval of one week between repeated tasks was most likely to lead to increased structural complexity and repair fluency but immediate repetition was most associated with increased speed fluency.

References

Ahmadian M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research, 15*(1), 35-59.

Ahmadian, M. J., Mansouri, S. A., & Ghominejad, S. (2017). Language learners' and teachers' perceptions of task repetition. *Elt Journal*. Amiryousefi, M. (2016). The differential effects of two types of task repetition on the complexity, accuracy, and fluency in computer-mediated L2 written production: a focus on computer anxiety. *Computer Assisted Language Learning, 29*(5), 1052-1068. doi:10.1080/09588221.2016.1170040

Azkarai, A., & García Mayo, M. d. P. (2017). Task repetition effects on L1 use in EFL child task-based interaction. *Language Teaching Research, 21*(4), 480-495.

Betz, F., Hothorn, T., & Westfall, P. (2010). *Multiple Comparisons Using R*. CRC Press.

Boers, F. (2014). A reappraisal of the 4/3/2 activity. *RELC Journal, 45*, 221–235.

Bui, G. (2014). Task readiness: Theoretical framework and empirical evidence from topic Familiarity, strategic Planning, and proficiency levels. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 63-94). Amsterdam, the Netherlands: John Benjamins. DOI: 10.1075/tblt.5.03gav

Bui, G., & Huang, Z. (2018). L2 fluency as influenced by content familiarity and planning: Performance, methodology and pedagogy. *Language Teaching Research*, *22*(1), 94-114. DOI: 10.1177/1362168816656650

Bui, G., & Skehan, P. (2018). Complexity, fluency and accuracy. In J. Liontas (Ed.), *TESOL encyclopedia of English language teaching*. Hoboken, NJ: John Wiley & Sons, Inc. DOI: 10.1002/9781118784235.eelt0046 Bui, G., & Teng, F. (2018). Exploring learners' behavioral patterns in two task-readiness conditions: A qualitative study. *Chinese Journal of Applied Linguistics, 41*(2), 129-149. Doi: 10.1515/cjal-2018-0008

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

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review, 24*(3), 369-378.

De Bot, K. (1992). 1992: A bilingual production model: Levelt" s speaking model adapted. Applied Linguistics 13, 1-24.

De Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning, 61*(2), 533-568.

Fukuta, J. (2016). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research, 20*(3), 321-340.

Gass, S., Mackey, A., Alvarez-Torres, M. J., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning, 49*(4), 549-581.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7*(2), 95-112.

Hong Kong Examinations and Assessment Authority. (2015). Benchmarking Study between IELTS and HKDSE English Language Examination. Retrieved from http://www.hkeaa.edu.hk/DocLibrary/IR/FAQs_IELTS_eng_Aug2013.pdf

Hsiu-Chen Hsu (2017) The combined effect of task repetition and post-task transcribing on L2 speaking complexity, accuracy, and fluency, T*he Language Learning Journal,* DOI: 10.1080/09571736.2016.1255773

Hunter, A-M. (2017). Fluency development in the ESL classroom : The impact of immediate task repetition and procedural repetition on learners' oral fluency. Unpublished PhD Thesis. University of Surrey.

Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System, 41*(3), 829-840.

Kormos, J. (2006). *Speech production and second language acquisition*: Routledge.

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in second language acquisition, 39*(1), 167-196.

Levelt, W. J. (1989). Speaking: From intention to articulation. ACL: MIT Press Series in Natural-Language Processing. MIT Press, Cambridge, Massachusetts.

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research, 4*(3), 221-250.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed). Mahwah, NJ: Lawrence Erlbaum Associates.

Meyers, L. S., Gamst, G.., & Guarino, A. J. (2006*). Applied multivariate research: Design and interpretatio*n. Thousand Oaks: Sage Publications.

Netten, J., & Germain, C. (2004). Theoretical and research foundations of intensive French. *Canadian Modern Language Review, 60*(3), 275-294.

Serrano, R. (2011). The time factor in EFL classroom practice. *Language Learning, 61*(1), 117-145.

Serrano, R., & Muñoz, C. (2007). Same hours, different time distribution: Any difference in EFL? *System, 35*(3), 305-321.

Skehan, P. (2009). Lexical performance by native and non-native speakers on language learning tasks. In H. Daller, D. Malvern, P. Meara, J. Milton, B. Richards, J. Treffers-Daller (Eds.). *Vocabulary studies in first and second language acquisition: The interface between theory and application*. London, UK: Palgrave Macmillan.

Skehan, P. (2014). *Processing perspectives on task performance*. Amsterdam, the Netherlands: John Benjamins.

Skehan, P., Foster, P., & Shum, S. (2016). Ladders and snakes in second language fluency. *International Review of Applied Linguistics*

Suzuki, Y., & DeKeyser, R. (2017). Effects of distributed practice on the proceduralization of morphology. *Language Teaching Research, 21*(2), 166-188.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Eds.), *Planning and task performance* (pp. 239-273). Amsterdam/Philadelphia: John Benjamins.

Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, accuracy and complexity. *TESOL Quarterly*, *50*(2), 369–393.

Wang, Z. (2014). On-line time pressure manipulations: L2 speaking performance under five types of planning and repetition conditions. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 27-62). Amsterdam, Netherlands: John Benjamins.

van de Guchte, M., Braaksma, M., Rijlaarsdam, G., & Bimmel, P. (2016). Focus on form through task repetition in TBLT. *Language Teaching Research, 20*(3), 300-320.

Appendix: the picture description task

Table 1: the effects of task repetition on CAF

Study	Spacing, participants, tasks, and research design	Increase in complexity	Increase in accuracy	Increase in fluency
Bygate (2001)	10 weeks; 48 language learners; between groups design; narrative task and interview	Yes	No	Yes
Hsu (2017)	1 week; 39 EFL learners; between-subject design; picture-based narrative tasks	No	No	No
Sample & Michel (2014)	1 week; 6 young participants; exploratory study; spot-the-difference task	No	No	Yes
Fukuta (2016)	1 week; 24 EFL learners; between groups design; narrative task	Not conclusive	Yes	Yes
Boers (2014)	Immediate; 10 ESL learners; counterbalance design; giving a talk on a topic;	Yes (after 2 repetitions)	Yes (after 2 repetitions)	Yes
Lynch & Maclean (2000; 2001)	Immediate; 14 participants from diverse backgrounds in an EAP setting; repeated measures design; poster carousal task	Yes (across 6 performances)	Yes (across 6 performances)	Yes
Thai & Boers (2016)	Immediate; 20 EFL learners; talking about one of their favourite movies; between groups design	Yes (after 2 repetitions)	Yes (after 2 repetitions)	Yes
Wang (2014)	Immediate; 77 EFL learners; counterbalanced design, oral narrative task	Yes	Yes	Yes
Hunter (2017)	Immediate	Yes (after 2 repetitions)	Yes (after 2 repetitions)	Yes

Table 2 Group-specific information of participants (SD in parentheses)

TR groups	N	Proficiency test mean	Age	Year of study
Immediate	14	17.39 (4.87)	18.86 (1.03)	1.43 (.51)
1-day interval	14	18.00 (3.71)	19.00 (1.11)	1.29 (.47)
3-day interval	14	17.36 (4.64)	19.07 (.92)	1.57 (.51)
1-week interval	15	16.93 (3.80)	18.80 (1.08)	1.33 (.49)
2-week interval	14	18.29 (4.51)	18.64 (.74)	1.50 (.52)
Total	71	17.58 (4.23)	18.87 (.97)	1.42 (.50)

Table 3 Dependent variables

Performance dimensions	Measures	Description
Structural complexity	No. of clauses per AS unit	The ratio of subordinate clauses per AS unit.
	Words per AS unit	The average number of words per AS unit.
Lexical complexity	Lexical diversity	The value of VocD obtained from the CLAN program by MacWhinney (2000)
Accuracy	Accurate clauses	The proportion of error-free clauses to all clauses
	Errors per 100 words	The number of errors in every pruned one hundred words.
Speed fluency	Speech rate	Total pruned words per minute after deletion of filled pauses, reformulations, replacements, false starts, and repetitions.
Breakdown fluency	No. of mid-clause pauses per 100 words	The average number of pauses greater than 0.4 seconds (following Skehan, 2014) per 100 pruned words.
	No. of filled pauses per 100 words	The average number of filled pauses, such as “er”, “hmm” and “um”, per 100 pruned words.
Repair fluency	Reformulations per 100 words	The average number of reformulations, operationalized as either phrases or clauses that are repeated with some modification to syntax, morphology, or word order, per 100 pruned words
	False starts per 100 words	The average number of false starts, operationalized as utterances that are abandoned before completion and that may or may not be followed by a reformulation, per 100 pruned words
	Repetitions per 100 words	The average number of repetitions, operationalized as words, phrases or clauses that are repeated with no modification whatsoever to syntax, morphology, or word order, per 100 pruned words
	Replacements per 100 words	The average number of replacements, operationalized as lexical items that are immediately substituted for another, per 100 pruned words.

Table 4 Wilkins' Lambda for Multivariate Tests

	Time		Time X Interval	
	F	Sig.	F	Sig.
Structural complexity	17.56	.000	.90	.46
Words per AS Unit	15.48	.000	.43	.78
Lexical diversity D	1.30	.25	.45	.76
Accurate clauses	.83	.36	.62	.64
Errors per 100 words	6.71	.01	.90	.46
Speech rate	7.99	.006	1.03	.39
No. of mi-clause pauses per 100 words	3.56	.06	1.14	.34
No. of filled pauses per 100 words	.44	.50	.38	.82
Reformulations per 100 words	1.11	.29	.53	.71
False starts per 100 words	.02	.87	.40	.30
Repetitions per 100 words	1.67	.20	.83	.51
Replacements per 100 words	2.66	.10	1.47	.22

Table 5: Effects of task repetition and spacing on complexity (SD in ())

		No interval	1-day interval	3-day interval	1-week interval	2-week interval	Sig.
No. of clauses per AS unit	Initial task	1.47 (.28)	1.38 (.27)	1.46 (.22)	1.42 (.24)	1.39 (.22)	Repetition: F = 17.56, p =. 00, η^2=.21 Interval: F = .82, p = .52, η^2=.04 Rep x Inter: F = .92, p =.46, η^2=.05
	Repeated task	1.56 (.26)	1.52 (.27)	1.63 (.24)	1.51 (.22)	1.42 (.18)	
Words per AS unit	Initial task	9.95 (1.53)	8.99 (1.65)	10.74 (2.02)	9.60 (1.63)	9.90 (1.30)	Repetition: F = 15.49, p =.00, η^2=.19 Interval: F = 2.39, p = . 06, η^2=.12 Rep x Inter: F = .44, p =.78, η^2=.02
	Repeated task	10.79 (2.35)	9.71 (1.84)	11.24 (1.34)	10.95 (1.56)	10.73 (1.38)	
Lexical diversity (D)	Initial task	41.19 (15.29)	38.16 (19.09)	45.57 (10.43)	51.55 (22.77)	33.94 (13.73)	Repetition: F = 1.31, p =.26, η^2=.01 Interval: F = 2.14, p =. 09, η^2=.11 Rep x Inter: F = .45, p =.77, η^2=.02
	Repeated task	45.94 (16.12)	42.79 (21.70)	46.29 (13.79)	50.31 (24.44)	35.07 (9.99)	

Table 6: Effects of task repetition and spacing on Accuracy (SD in ())

		No interval	1-day interval	3-day interval	1-week interval	2-week interval	Sig.
Accurate clauses	Initial task	.50 (.17)	.51 (.13)	.43 (.21)	.49 (.12)	.47 (.14)	Repetition: F = .84, p =. 36, η^2=.01 Interval: F = .99, p = .42, η^2=.05 Rep x Inter: F = .63, p =.64, η^2=.03
	Repeated task	.55 (.15)	.53 (.18)	.46 (.17)	.49 (.09)	.44 (.11)	
Errors per 100 words	Initial task	10.03 (4.44)	10.66 (3.37)	11.43 (3.76)	11.18 (4.01)	10.96 (3.94)	Repetition: F = 6.72, p =.01, η2=.09 Interval: F = 1.30, p = .28, η^2=.02 Rep x Inter: F = .91, p =.47, η^2=.05
	Repeated task	7.50 (3.04)	9.49 (4.04)	10.32 (3.44)	10.04 (2.42)	11.13 (2.59)	

Table 7: Effects of task repetition and spacing on speed and breakdown fluency (SD in ())

		No interval	1-day interval	3-day interval	1-week interval	2-week interval	Sig.
Speech rate	Initial task	80.79 (21.00)	101.73 (24.79)	88.46 (20.92)	103.10 (24.99)	108.38 (15.98)	Repetition: $F = 8.00$, $p = .01$, $\eta^2 = .11$
	Repeated task	92.79 (22.31)	110.07 (30.39)	93.98 (20.80)	105.86 (29.10)	108.54 (14.25)	Interval: $F = 2.90$, $p = .03$, $\eta^2 = .15$
							Rep x Inter: $F = 1.03$, $p = .40$, $\eta^2 = .05$
No. of mi-clause pauses per 100 words	Initial task	14.66 (6.44)	12.32 (6.36)	14.80 (6.68)	12.07 (4.99)	9.97 (3.52)	Repetition: $F = 3.56$, $p = .06$, $\eta^2 = .05$
	Repeated task	11.82 (7.68)	11.77 (6.15)	11.71 (6.76)	12.22 (9.70)	10.04 (3.89)	Interval: $F = .75$, $p = .56$, $\eta^2 = .04$
							Rep x Inter: $F = 1.15$, $p = .34$, $\eta^2 = .06$
No. of filled pauses per 100 words	Initial task	7.87 (6.85)	8.13 (6.62)	12.02 (7.61)	1.88 (1.78)	5.61 (4.66)	Repetition: $F = .45$, $p = .51$, $\eta^2 = .007$
	Repeated task	7.54 (7.09)	8.21 (5.86)	13.86 (9.71)	1.42 (2.83)	7.07 (7.04)	Interval: $F = 7.92$, $p = .00$, $\eta 2 = .32$
							Rep x Inter: $F = .38$, $p = .82$, $\eta^2 = .02$

Table 8: Effects of task repetition and spacing on repair fluency (SD in ())

		No interval	1-day interval	3-day interval	1-week interval	2-week interval	Sig.
Reformulations per 100 words	Initial task	3.36 (2.05)	2.65 (2.16)	2.90 (2.33)	2.57 (1.83)	3.09 (2.92)	Repetition: $F = 1.12$, $p = .29$, $\eta^2 = .01$ Interval: $F = 1.06$, $p = .37$, $\eta^2 = .06$ Rep x Inter: $F = .53$, $p = .74$, $\eta^2 = .03$
	Repeated task	3.39 (2.41)	2.16 (1.35)	3.00 (2.62)	1.60 (1.49)	2.89 (2.12)	
False starts per 100 words	Initial task	.56 (.66)	.40 (.66)	.59 (.61)	.35 (.66)	.35 (.55)	Repetition: $F = .03$, $p = .87$, $\eta^2 = .00$ Interval: $F = .58$, $p = .68$, $\eta^2 = .03$ Rep x Inter: $F = .40$, $p = .81$, $\eta^2 = .02$
	Repeated task	.26 (.43)	.40 (.63)	.63 (.72)	.45 (.98)	.42 (.63)	
Repetitions per 100 words	Initial task	1.99 (2.08)	2.99 (2.04)	3.98 (5.82)	1.35 (1.16)	2.05 (1.69)	Repetition: $F = 2.67$, $p = .11$, $\eta^2 = .02$ Interval: $F = 2.51$, $p = .05$, $\eta^2 = .12$ Rep x Inter: $F = 1.47$, $p = .22$ $\eta^2 = .04$
	Repeated task	2.63 (2.54)	3.02 (1.55)	5.08 (6.22)	.89 (.99)	2.56 (2.02)	
Replacements per 100 words	Initial task	.63 (.74)	.44 (.66)	.94 (1.43)	.17 (.30)	.40 (.62)	Repetition: $F = 1.68$, $p = .20$, $\eta^2 = .04$ Interval: $F = 2.15$, $p = .09$, $\eta^2 = .01$ Rep x Inter: $F = .83$, $p = .51$, $\eta^2 = .01$
	Repeated task	.46 (1.06)	.97 (1.35)	1.2 (1.58)	.30 (.52)	.47 (.61)	

Table 9: The spacing conditions most conducive to measures affected

Measures	Interval with the strongest effect
Complexity: Words per AS unit	One week
Fluency: Words per minute	Immediate repetition (no interval)
Fluency: No. filled pauses per 100 words	One week
Fluency: Repetition per 100 words	One week