Task repetition and second language speech processing

Authors:

Craig Lambert
Curtin University, School of Education
Kent Street, Bentley
Perth, West Australia 6845, Australia
craig.lambert@curtin.edu.au

Judit Kormos
Lancaster University
Department of Linguistics and English Language
Lancaster, United Kingdom
LA1 4YL
j.kormos@lancaster.ac.uk

Danny Minn
Kitakyushu University, Center for Fundamental Education,
Kokura Minami-ku Kitagata 4-2-1
Kitakyushu-shi, Fukuoka-ken 802-8577, Japan
<danminn@kitakyu-u.ac.jp>

TO BE PUBLISHED IN STUDIES IN SECOND LANGUAGE ACQUISITION

Author’s accepted version

Abstract

This study examines the relationship between the repetition of oral monologue tasks and immediate gains in L2 fluency. It considers the effect of aural-oral task repetition on speech rate, frequency of clause-final and mid-clause filled pauses, and overt self-repairs across different task types and proficiency levels and relates these findings to specific stages of L2 speech production (conceptualization, formulation and monitoring). Thirty-two Japanese learners of English sampled at three levels of proficiency completed three oral communication tasks (instruction, narration and opinion) six times. Results revealed that immediate aural-oral same task repetition was related to gains in oral fluency regardless of proficiency level or task type. Overall gains in speech rate were the largest across the first three performances of each task type, but continued until the fifth performance. More specifically, however, clause-final pauses decreased until the second performance, mid-clause pauses to the fourth, and self-repairs decreased only after the fourth performance indicating that task repetition may have been differentially related to specific stages in the speech production process.

BACKGROUND AND MOTIVATION

A primary issue for research on incidental SLA is how to optimize learners' ability to use language in conjunction with communicative task performance. Definitions of such tasks have varied, but for the purpose of the present study, the term "task" will be used, following Ellis (2009), to refer to L2 learning activities that meet four criteria:

1. There is a primary focus on meaning,
2. There is a gap which necessitates communication,

3. Learners must draw on their own resources to complete it,

4. There is a communicative outcome beyond the use of language for its own sake.

When these criteria are met, tasks can be argued to play an important role in L2 pedagogy in that they provide learners with opportunities to employ their linguistic repertoire under relatively natural conditions.

One factor in the implementation of such tasks that has been argued to play a crucial role in optimizing language use and that has received considerable theoretical and empirical attention in the L2 literature is the planning processes that learners engage in in conjunction with the performance of these tasks (see Pang & Skehan, 2014, for a recent overview). Ellis (2005) identifies two points in the pedagogic process at which opportunities for planning might be manipulated. The first is before the task begins. At this stage, learners might either be allowed time to plan the language or content that they will use during the task (either in the L1 or in the L2) or given the opportunity to rehearse their performance of the task. In the latter case, they complete a sequence of tasks in which they either repeat the exact same task or a parallel version of the task with slightly different content (Bygate, 2001). The second point at which learners' opportunity to plan their performance might be controlled is during the task performance itself. Learners can either be allowed unlimited time to complete the task, or time limits might be imposed on their performances (Maurice, 1983; Nation, 1989).

In line with the definition of task outlined above, however, task repetition might be argued to necessitate a different interlocutor each time the task is repeated. A change in interlocutor preserves task integrity by requiring learners to create original meanings each time the task set is performed. In the present study, for example, participants performed monologue tasks in pairs alternating as speaker and listener with a different interlocutor each time as they might if they were repeating the tasks in a classroom or real-world setting. This approach could be argued to add to the external validity of the repetition process in that the exposure that learners receive involves both input-based and output-based versions of the task. The construct of task repetition as it is operationalized in the present study thus represents a combination of task repetition and interlocutor input. It is thus referred to as *aural-oral* task repetition.

Task repetition of this sort might help learners improve their performances by allowing them to activate, refine, and optimize their linguistic resources for the purpose of successfully completing a given task in line with the specific communicative demands that it entails. Empirical studies on task repetition have asked learners to repeatedly engage in exactly the same task (e.g. Ahmadian & Tavakoli, 2011; Bygate, 1996, 1999, 2001), in the same type of task with slightly different content (Gass, Mackey, Alvarez-Torres & Fernandez-Garcia, 1999), or in a task that has the same communicative goal but different content and interlocutors each time (Lynch & Maclean, 2000, 2001). Several empirical studies have also examined the effects of repetition in conjunction with decreasing the time available on consecutive performances (e.g., de Jong & Perfetti, 2011; Nation, 1989; Arevart & Nation, 1991, 1993). Furthermore, repetition effects have been investigated immediately (e.g. Lynch & Maclean, 2000, 2001; Wang, 2014) or after days (Gass et al., 1999) or weeks (e.g. Ahmadian & Tavakoli, 2011; Bygate, 1996, 1999, 2001). Finally, speakers have repeated their performance once (e.g. Bygate, 1996, 1999, 2001; Wang, 2014) or several times ranging from 3 or 4 (Gass et al., 1999) to 11 (Ahmadian, 2011).

Of the different forms of planning proposed by Ellis (2005), rehearsal (or task repetition) has been shown to have the most robust effects on L2 fluency (Ahmadian &

Tavakoli, 2011; Arevart & Nation, 1991, 1993; Bygate, 2001; Lynch & Maclean, 2000; Nation, 1989; Wang, 2014). Fluency is generally defined as skilled L2 performance, referring to rapid, smooth and accurate communication of one's intentions during on-line processing (Lennon, 2000, p. 26). Fluent performance thus entails the efficient functioning of speech production processes under the constraints of real-time oral interaction.

The modular model of L1 speech production proposed by Levelt (1989, 1999) and adapted to L2 speech production by de Bot (1992) and Kormos (2006) postulates three primary stages of speech production. The first is a conceptualization stage in which the speaker selects information from world knowledge to include in a message and organizes it into an information structure to create a pre-verbal plan. The second is a complex set of procedures referred to as a formulation stage in which the pre-verbal plan is encoded grammatically and phonetically. Concepts and their relational structure are argued to be projected onto a phrase structure driven by lemmas drawn from the speaker's mental lexicon that have associated semantic, syntactic, morphological, and phonological properties. This phrase structure is then encoded with phonetic and prosodic information to produce a phonetic plan. The third and final stage of speech production is then articulation in which the phonetic plan is buffered and parsed as syllables at the motor level.

The essence of this model, as it relates to L2 fluency, is that these three stages are assumed to operate in parallel. Although each module is expected to work on specific input and generate specific output for the next module, they are hypothesized to operate simultaneously provided that processing in the parallel module is sufficiently automatic (Levelt, 1989, 1999). For proficient speakers, formulation may be largely automatic and allow for parallel processing with other modules, whereas conceptualization and monitoring may both require attention and thus rely on serial processing. For lower proficiency speakers, however, lexical retrieval and grammatical encoding might also require attentional resources and hence might result in varying degrees of breakdown in parallel processing (de Bot, 1992; Kormos, 2006). L2 learners thus face a number of challenges in the processes leading from conceptualization to articulation, one of the most important being limitations in attention and working memory capacity which are essential in managing speech production (Kormos, 2011). The fluency of L2 performance reflects the efficient functioning of speech production mechanisms including the automaticity of encoding processes, the conceptual demands in planning the content of the message, and the skillful handling of performance breakdowns. This aspect of fluency is referred to as cognitive fluency by Segalowitz (2010).

Cognitive fluency can be inferred from the analysis of *utterance fluency*, which designates the temporal variables of speech or the "oral features of utterances that reflect the operation of underlying cognitive processes" (Segalowitz, 2010, p. 48). In task-based studies, utterance fluency is often argued to consist of three sub-constructs: breakdown fluency, repair fluency, and speed fluency (for recent discussions, see Bosker, Pinget, Quené, Sanders, & De Jong, 2013; De Jong, Steinel, Florijn, Schoonen & Hulstijn, 2012). *Breakdown fluency* is generally considered to be most accurately measured by pause frequency, *repair fluency* is usually operationalized as the frequency of overt repairs or reformulations in L2 learners' speech (see De Jong et al., 2012; Götz, 2013), and *speed fluency* is typically assessed by speech rate or the number of pruned syllables uttered per second (Ellis & Barkhuizen, 2005). Although it is difficult to relate these fluency measures to a specific speech production mechanism, it can be argued that filled and unfilled pauses *between* clauses are indicators of conceptualization and content planning whereas filled and unfilled pauses *within* clauses signal breakdowns in lexical and syntactic encoding (Butterworth, 1975), and the frequency of overt self-repairs might be argued to reflect the availability of attentional resources for monitoring (Kormos, 1999). Finally, speech rate is an overall measure of the speed with which an utterance is produced, but it is not a pure indicator of encoding speed as it includes hesitation time (see Götz, 2013).

Particularly relevant to the present study is Wang's (2014) research which compared various approaches to providing learners with opportunities to plan and found much larger effect sizes for repeating a task once than for other types of planning (see also Skehan, Xiaoyue, Quian & Wang, 2012). However, in Wang's (2014) study, speech rate increased and the average length of unfilled pauses at clause boundaries decreased, whereas the average length of unfilled pauses within clauses and the frequency of reformulations were unaffected by repetition. The question arises of why this might have been the case and whether subsequent repetitions of the task might not have improved other aspects of learners' L2 fluency. Wang's findings suggest that one task repetition enhances the speed with which students can deliver their message as well as reduces the time needed to pause at clause boundaries for generating ideas. Nevertheless, subsequent repetitions may be required before learners can acquire sufficient access to L2 resources (lexis and syntax) to avoid breakdowns in speech processing as reflected in the frequency of mid-clause pauses and devote attention to monitoring their output more carefully as reflected in the frequency of self-repairs and reformulations.

Based on Levelt's (1989, 1999) speech production model, Bygate (2001) argues that when L2 learners perform a task for the first time, their speech production system needs to execute all the relevant processing steps under time pressure. Hence, based on the limited attentional model of speech production (Kormos, 2006; Skehan, 2009), it could be argued that upon the first performance of a given task learners have to distribute their attentional resources strategically between conceptualization, formulation, and monitoring. They also have to handle performance breakdowns due to incomplete lexical or syntactic knowledge representation or slow speed of access to these representations (Dörnyei & Kormos, 1998). Task repetition allows L2 learners to rely on previously conceptualized task content and to activate recently used linguistic constructions to express their message. This might reduce the attentional demands on learners to conceptualize, encode and monitor their messages simultaneously.

The effects of task repetition, however, might vary depending on the speech production demands of tasks. Communicative tasks can differ in terms of the load they pose on working memory in various stages of speech production (Kormos, 2006; Skehan, 2009). Opinion tasks, for example, might require increased attention to conceptual planning, either at the expense of attention to linguistic encoding or in addition to it. By contrast, tasks in which learners have to convey familiar information, or information provided visually, such as personal or picture-based narrations or instructions, might pose lower attentional demands during conceptualization and allow learners to reallocate attentional resources to linguistic encoding. In studies of task repetition, it is therefore important to examine the extent to which the hypothesized benefits of task repetition vary across tasks of differing demands.

The effects of task repetition on speech processing might also vary depending on the proficiency level of the speakers who complete the tasks. As the automaticity of linguistic encoding is strongly associated with L2 proficiency (Segalowitz, 2010), higher proficiency L2 speakers might have more attentional resources available for conceptualizing task content. For this reason, task repetition might result in fewer gains for more advanced learners as one of its primary functions is to allow learners to activate relevant content and linguistic encoding processes. Furthermore, language proficiency may interact with task demands so that higher proficiency learners, for example, would improve as a result of repeating tasks that are high in conceptualization demands, but not on those in which conceptual planning requires less attention. In addition to task demands, it is thus necessary to consider the possible role of proficiency level in moderating the effects of task repetition. To our knowledge, no previous research has investigated how task characteristics and learner proficiency, independently and in interaction with each other, affect fluency gains in repeated task performance.

THE STUDY

It could thus be argued, on the one hand, that performing tasks multiple times within a relatively short time interval is beneficial to the development of learners' speech processing ability. Most empirical work, however, has only investigated learners' performance over a small number of task repetitions (for exceptions see Ahmadian, 2011, Lynch & McLean, 2000, 2001). To date, little is known about how many repetitions are actually needed for learners to optimize the different aspects of their L2 fluency in the short term and how this might vary with task type and learner proficiency level. On the other hand, task repetition might also result in boredom and fatigue (Bygate, 2001). It is thus also important to investigate learners' reactions to repeating tasks and the value that they perceive in doing it. The present study attempts to address both of these issues by considering language production together with the reactions of 32 Japanese learners of English at three proficiency levels to performing three types of tasks six times each within the time frame of a typical L2 lesson.

Research Questions

1. How does repeating the same task as speaker and listener (i.e., aural-oral task repetition) with different interlocutors affect L2 fluency in the short term?
2. Do the demands of different tasks moderate the effects of aural-oral task repetition on L2 fluency?
3. Do speakers' proficiency levels moderate the effects of aural-oral task repetition on L2 fluency?
4. What are the learners' perceived value of aural-oral task repetition, the optimum number of repetition times, as well as the extent to which repetition brings about boredom or fatigue?

METHODS

Design

The study employed a three-way repeated-measures design. The first within-subjects factor was task repetition at six levels, and the second was task type at three levels (instruction, narration, and opinion). The third between-subjects factor was proficiency level which was established based on band levels of the TOEIC test (Educational Testing Service, 2008). The four dependent variables in the study were: (1) speech rate, operationalized as the number of syllables per second based on pruned transcriptions of oral discourse; (2) clause-final pausing, operationalized as the ratio of filled pauses between clause boundaries to the total number of pruned syllables produced; (3) mid-clause pausing, operationalized as the ratio of filled pauses within clause boundaries to the total number of pruned syllables produced; and (4) overt self-repairs, operationalized as the total number of self-corrections or rephrasing to the total number of pruned syllables produced. Post-performance questionnaires were also used to triangulate the findings for speech production, gain deeper insight into the range of variables that might be involved in task repetition, and better understand the responses of learners at different levels of proficiency to repeating the tasks used in the study.

Participants

Thirty-two Japanese English learners, ranging from 18 to 23 years of age, at a large public university in southern Japan participated in the study. These learners were sampled from the entire population of English learners at the university at three levels of proficiency based on their TOEIC scores (Educational Testing Service, 2008). The high-level group was roughly equivalent to B2 and C1 levels on the Common European Framework of Reference (CEFR) (Council of Europe, 2001); the mid-level group to B1 level on the CEFR; and the low-level learners to A2 level on the CEFR. The participants' TOEIC scores were approximately normally distributed both within and across the three proficiency bands (see Table 1).

Insert Table 1 around here

Materials

Based on Yule's (1997) typology of referential communication tasks, three tasks differing in discourse genre (instruction, narration, and opinion) were used in the study. The instruction task presented participants with a traditional problem in which a farmer has to get a fox, a chicken and some wheat across a river, but he has a boat big enough only for himself and one other thing, and he cannot leave the fox and the chicken alone together, nor can he leave the chicken and the wheat alone together. It also provided an eight-picture sequence which illustrated how to solve the problem and asked them to explain the solution six times to different interlocutors. The narration task was based on a four-frame picture story taken from the Pre-1 Level of the EIKEN English Test (Nihon Eigo Kentei Kyokai, 2009) which involved a group of shop owners arriving at a solution to a local graffiti problem and asked participants to narrate the story to these six interlocutors. Finally, the opinion task presented them with two photographs from a family trip to a zoo and asked them to give their opinions on what they saw in the pictures, what they thought was happening, who they thought took them, and why they thought they were taken to the six interlocutors. In each case, participants were given a one-sentence prompt with which to begin in order to establish the respective discourse frames.

In addition to the inherent differences in discourse demands between the instruction, narration, and opinion genres (for discussions see Berman, 2008; Yule, 1997), the format of the tasks could also be argued to differ in terms of the amount of conceptual and linguistic structure that they provided. The instruction and narration tasks had a clear sequential information structure. In the instruction task, learners were given pictures that supplied the specific content to complete the task, and each successive step allowed them to recycle the same lexis and sentence frames. The narration task was similar in that it provided learners with the specific content to be related, but it allowed far less repetition of language from one frame of the story to the next. Finally, the opinion task required learners to supply the content required to complete the task themselves (i.e., what they saw in the pictures, what they thought was happening, who they thought took them, and why they thought they were taken). In other words, the opinion task did not provide a clear information structure in the way that the instruction and narration tasks did (see Tavakoli & Foster, 2008, on the effects of content structure on L2 speech performance). The open nature of the opinion task might also be expected to result in less sense of a clear outcome.

Finally, two questionnaires were used in the study. The first was a brief background questionnaire which focused on English language experience, and the second was a post-performance questionnaire on learners' thoughts on the value of repeating each of the tasks. For each of the three tasks, participants were asked: (1) whether and in what ways they felt their performance improved as a result of repetition, (2) whether their performance improved through the sixth performance and, if not, how many repetitions were enough, and (3) whether and in what ways it was useful to complete the task with multiple partners.

Procedures

Advertisements were circulated to all English learners at the university offering payment in cash for participation in a two-hour English conversation research project. Participants were accepted in the order in which they applied and allocated to groups based on their proficiency level and availability. They all received information about the aims of the research and signed a consent form.

In addition to the three monologue tasks reported in the present study (instruction, narration, and opinion), a dialogue opinion task was also included in a balanced Latin square design when the data was collected. The order in which these four tasks were performed was counterbalanced into four task sets which were then performed six times each. In other words, the 32 participants were allocated to one of four groups and each group of eight performed one task set six times during a separate recording session.

Each group of eight participants then worked in pairs. They performed each task in the task set twice (once as the speaker and once as the listener). Speaker and listener roles were alternated after each task performance so that each task in the set was completed by both participants before they moved on to the next one. When participants had finished all four tasks with one partner, they switched partners and repeated the same process with another partner until they had performed the task set six times with different partners each time. Participant pairing was controlled across the design so that each participant worked with partners from each of the three proficiency levels. To accomplish this, each of the eight participants in a given session was given a letter or a number (A-B-C-D-1-2-3-4). They were then systematically alternated in terms of partner and speaker-listener order across the six performances of their task set. Table 2 indicates the specific pairings of learners and the order of speakers used in the study.

Insert Table 2 around here

It can be seen in Table 2 that proficiency level was generally distributed across the design. Each participant worked with six of the seven other participants in the study, but the order in which they worked with speakers of different levels varied across the design. Furthermore, the order in which the participants performed initially in the role of speaker or listener in each pair was alternated across the design so that each participant initially functioned as the speaker or listener a comparable number of times across the design as a whole. However, interlocutor factors (proficiency and order) could not be counterbalanced perfectly. Even if they had, interlocutor effect would only have been distributed across the design rather than controlled. The design used in the study thus combines the effect of task repetition with that of interlocutor input. The advantages and disadvantages of the design will be considered in the *Discussion and Conclusions* section of the paper.

Finally, participants were given 120 seconds to complete each task for the first and second repetitions. Performance time was reduced to 90 seconds for the remaining four repetitions as piloting in classes with similar learners had shown this time was adequate for completing the tasks at their own pace. This was confirmed in the performances recorded for the project. Most participants had finished slightly before the time had expired. Thus, the reduction in time after the second performance in the sequence cannot be argued to have constituted a pressured performance, reducing or eliminating during-task planning time, as participants generally did not use all of the time that they had available.

Each participant thus performed the three monologue tasks in the role of the speaker 18 times (six performances of the four task types) and in the role of the listener 18 times within a period of approximately 90 minutes or the length of a typical university-level English class.

Analysis

The study is based on 576 task performances constituting approximately 16 hours of oral data. Each of the 32 participants produced approximately 30 minutes of speech (18 task performances). These performances were transcribed into AS-units following the procedures of Foster, Tonkyn, and Wigglesworth (2000). Speech rate was then calculated as the number of pruned syllables per second from the first syllable after the prompt sentence to the end the task performance (Ellis & Barkhuizen, 2005).

As automated analysis of unfilled pause data proved impossible due to background noise in the classroom, the frequency of *filled pauses* was counted manually. Filled pauses were identified as “non-lexical fillers such as *er* and *uhm* and elongations of sounds (drawls)” (Gut 2009, p. 80). The ratio of *clause-final* and *mid-clause* pauses to total pruned syllables was then calculated for each performance.

The number of overt self-repairs during each performance was also counted manually. Overt self-repairs were defined as self-initiated, self-completed corrections when the speaker halts the speech flow and executes a modification (Kormos, 1999, p.

313). [1] The ratio to total pruned syllables was then calculated (see Appendix for an example of the data analysis).

Of the 576 task performances, 36 (6.25%) were selected across tasks and speakers and double coded for pause and self-repair frequencies by the second author and a research assistant. Two-tailed Pearson correlations revealed inter-coder reliability to be very high for clause final pauses (r =.956, N=36, p <.0001), mid-clause pauses (r =.952, N=36, p <.0001), and self-repairs (r =.966, N=36, p <.0001). The research assistant then coded the remainder of the performances.

Preliminary screening of the data revealed some positively skewed score distributions, particularly in the case of the two pause measures. This was corrected with square root transformation. Following transformation, a comparison of skew and kurtosis values with their standard errors revealed no z-scores higher than +/-1.96, and the data were thus assumed to be approximately normally distributed for the purpose of the statistical analyses used in the study. In addition, Levene's tests for each distribution revealed acceptable homogeneity of variance (p < .05) on each of the variables in the study. Multivariate analysis of variance (MANOVA) was then conducted using SPSS22 for Windows followed by post-hoc univariate analyses and pairwise comparisons with Fischer's LSD tests to test the null hypotheses that there were no statistically significant differences in speech rate, clause-final pausing, mid-clause pausing, or self-repair between: (1) the six levels of repetition; (2) the three task types; (3) the three proficiency levels, and (4) that there were no significant interactions between these variables individually or in combination. Cohen's d (Cohen, 1988) was then calculated to determine the effect sizes of the relevant pairwise comparisons. Cohen recommends interpreting effect sizes above d = 0.2 as small, above d = 0.5 as medium, and above d =

0.8 as large, and these benchmarks have provided the standard for interpreting effect sizes in social science research. However, based on a comprehensive survey of research in the field of SLA, Plonsky and Oswald (2014) argue that Cohen's scale underestimates the range of effects typically obtained in L2 research, and recommend field-specific benchmarks of small (d = .40), medium (d = .70), and large (d = 1.00) in interpreting effect sizes in SLA research. We have adopted the latter criteria for interpreting the effect sizes of the pairwise comparisons in the present paper.

RESULTS

Using Pillai's trace, multivariate analysis of variance revealed a significant main effect and a very large effect size for aural-oral task repetition on L2 speech fluency, V = 0.957, F (20, 10) = 11.163, p = .000, partial η^2 = .957. Cohen (1988, see also Tabachnick & Fidell, 2007), for example, recommends using .01, .09 and .25 in interpreting small, medium and large effects of partial η^2, respectively. Furthermore, this analysis revealed no significant interactions between this form of repetition, task type, and proficiency level individually or in combination with small effect sizes in all of these comparisons: (1) repetition*proficiency, V = 0.280, F (40, 580) = 1.092, p = .326, partial η^2 = .070, (2) repetition*task, V = 0.150, F (40, 1160) = 1.130, p = .268, partial η^2 = .038, and (3) repetition*proficiency*task, V = 0.244, F (80, 1160) = .940, p = .627, partial η^2 = .061. This indicates that the overall effects of aural-oral task repetition were generally robust for speakers of different proficiency levels completing tasks of differing discourse demands. Finally, follow-up univariate analyses revealed that this repetition effect was statistically significant across all four aspects of L2 fluency, but with a very large effect size for speech rate, and medium effect sizes for the other pause and self-repair phenomena observed: (1) speech rate, $F = 122.092$, $p = .000$, partial $\eta^2 = .808$, (2) clause-final filled pauses, $F = 6.306$, $p = .000$, partial $\eta^2 = .179$, (3) mid-clause filled pauses, $F = 9.243$, $p = .000$, partial $\eta^2 = .242$, and (4) self-repair, $F = 3.072$, $p = .019$, partial $\eta^2 = .096$. In the following sections, pair-wise comparisons with Fischer's LSD tests are discussed in order to elucidate the specific nature of the repetition effect on these four aspects of L2 fluency.

Speech Rate

Pairwise comparisons revealed that the aural-oral repetition effect on the rate of L2 speech production was due to differences between all performances (1-2 $p = .000$, $d = 1.53$; 2-3 $p = .000$, $d = 1.48$; 3-4 $p = .016$, $d = 0.48$; 4-5 $p = .000$, $d = 0.84$) except Performance 5 and 6 which was not significant ($p = .943$, $d = 0.05$). In other words, the effect of immediate aural-oral task repetition as speaker and listener with different interlocutors on speech rate continued to increase significantly through the fifth performance of the tasks in the study. However, the effect size was large according to the criteria posited by Plonsky and Oswald (2014) across the first three performances only and small to medium thereafter through the fifth performance (see Figure 1 and Table 3).

Insert Figure 1 around here

Not only did gains in speech rate begin to level off after the third repetition, but the pattern of development also began to diverge across the task types with the opinion task showing a less linear pattern of development than either the instruction or the narration task (see Figure 2 and Table 3).

Insert Figure 2 around here

Furthermore, the pattern observed for this form of task repetition on speech rate was very consistent across all three proficiency levels (see Figure 3 and Table 3).

Insert Figure 3 and Table 3 around here

The results thus indicate a robust connection between task repetition and speech rate and point to the possibility of a ceiling effect for immediate practice on speech rate at five repetitions.

Clause-Final Pausing

Pairwise comparisons revealed that the frequency of pauses at clause boundaries decreased significantly from the first to the second performance only (p = .047, d = 0.40) (cf., Wang, 2014), but the differences between subsequent repetitions did not reach statistical significance at the .05 level (2-3 p = .146, d = 0.21; 3-4 p = .071, d = 0.36; 4-5 p = .921, d = 0.03; 5-6 p = .576, d = 0.08). The effect of aural-oral task repetition on clause-final pausing was relatively small according Plonsky and Oswald (2014) and limited to the first two task performances (see Figure 4 and Table 4).

Insert Figure 4 around here

Furthermore, there was considerable variation in clause-final pausing across tasks. The largest gains were between the second and third repetition for the opinion tasks, between the first and second repetition of the narration task, and negligible on all repetitions the instruction task (see Figure 5 and Table 5). However, we see a generally descending pattern until the fourth performance on all task types after which non-linear patterns of development begin to emerge (see Figure 5).

Insert Figure 5 around here

Finally, comparison of clause-final pauses across the three proficiency levels revealed that the two higher proficiency groups showed a consistent pattern of decreased pausing across their first four performances of the task set (see Figure 6 and Table 4). This pattern is consistent with the pattern in Figure 4. The lowest proficiency speakers, on the other hand, showed fluctuating progress after their second performance of the task set (see Figure 6 and Table 4).

Insert Table 4 and Figure 6 around here

Mid-Clause Pausing

Post-hoc pairwise comparisons showed weak effects for aural-oral task repetition on mid-clause filled pauses. None of the differences between any two consecutive performances were statistical significant at the .05 level (1-2 p = .095, d = 0.52; 2-3 p = .110, d = 0.34; 3-4 p = .272, d = 0.13; 4-5 p = .307, d = 0.14; 5-6 p = .840, d = 0.04). Although the effect size for the first repetition was larger than the others, it was still small according to the benchmarks proposed by Plonsky and Oswald (2014). On the other hand, the difference between every second performance up to the fifth performance did reach statistical significance (1-3 p = .001, d = .71; 2-4 p = .003, d = .50; 3-5 p = .049, d = .30), whereas the difference between Performances 4-6 did not (p = .595, d = .12). Thus, the effects of aural-oral task repetition on mid-clause pausing did not dissipate until the fourth performance but the effect size decreased from medium to small (Plonsky and Oswald, 2014) (see Figure 7 and Table 5).

Insert Figure 7 around here

Nevertheless, the effects observed in Figure 7 were generally representative of participants' performance on all three of the tasks (see Figure 8 and Table 5). With the exception of the second performance of the narration task, we see a clearly descending pattern through the third performance on all task types after which mid-clause pausing behavior begin to fluctuate (see Figure 8).

Insert Figure 8 around here

Furthermore, comparison of the performance of the three proficiency levels reveals that the initial effect of aural-oral repetition on mid-clause filled pauses was most pronounced between the lowest proficiency level learners' first and the second performances of the task set after which all groups show a generally descending pattern in mid-clause pausing until the fourth performance (see Figure 9 and Table 5).

Insert Figure 9 and Table 5 around here

Self-Repair

Finally, pairwise comparisons revealed that the frequency of self-repairs and reformulations followed a pattern that was quite distinct from either pause phenomenon. None of the differences between any combinations of the first four performances reached statistical significance at the .05 level (1-2 p = .922, d = .02; 2-3 p = .356, d = .14; 3-4 p = .614, d = .16; 4-5 p = .216, d = .14; 5-6 p = .980, d = .08). However, the differences between the fifth repetition and the first two performances 1-5 (p = .017, d = 0.45), 2-5 (p = .01, d = 0.48), and the differences between the sixth performance and the first three performances were significant: 1-6 (p = .017, d = 0.50), 2-6 (p = .015, d = 0.51), 3-6 (p = .020, d = 0.49) although the effect sizes were small according to Plonsky and Oswald (2014). Thus, significant gains in L2 fluency due to decreased self-repair (see Figure 10 and Table 6) did not begin until decreases in both clause-final and mid-clause pausing had dissipated.

Insert Figure 10 around here

Likewise, when participants' performance on each task is compared, production on the instruction and narration tasks followed a similar pattern throughout the first five repetitions, but their performance on the opinion task showed a sudden decrease in fluency due to an increase in self-repair during Performance 4 (see Figure 11).

Insert Figure 11 around here

However, there was considerable variability between proficiency levels with respect to self-repair (see Figure 12 and Table 6). The frequency of self-repairs is stable between the first two performances of the task set for all groups when clause-final pausing effects were significant (see Figure 4), and relatively stable through the third performance when mid-clause pausing effects were the largest (see Figure 7). However, there is considerable fluctuation in self-repair behavior between groups in the latter performances of the task set when significant differences in self-repair behavior began to emerge, indicating that self-repair may be a fluency variable that is dependent to some extent on proficiency.

Insert Figure 12 and Table 6 around here

Speaker Perceptions

Participants' impressions were elicited regarding: (a) the value of repeating each task type, (b) the optimal number of task repetitions for each task type, and (c) the value of repeating tasks with different partners. Table 7 summarizes the results of this analysis.

Insert Table 7 around here

Overall, participants felt repetition to be helpful. Nearly all of them reported that repeating the task was useful on the opinion and narration tasks, and three-quarters of them indicated that it was beneficial on the instruction task. The majority of their comments fell into two categories. Approximately 40% mentioned that repetition improved their fluency on the tasks. Furthermore, between a quarter and half of the learners (24% instruction; 35% narration; 48% opinion) indicated that repetition assisted them in recalling useful words and expressions or that they were able to learn useful words and expressions from their partners. It is interesting to note that usefulness in terms of incorporation was felt to be the most relevant for the opinion task where task content was open and lowest for the instruction task where the task was structured and relatively inflexible with the narration task falling in between.

Other comments did not fit into these two categories, but provided information on participants' perceptions on the tasks and other situational variables. One participant, for example, said that speaking with a high proficiency interlocutor helped her improve her English speaking skills. Only three of the 32 participants (9.4%) expressed boredom after the fourth performance. On the opinion and instruction tasks, all three were in the high-level group. On the narration task, one participant from each group expressed boredom. In addition, a few participants recalled that during the later repetitions, they stopped incorporating new linguistic constructions into their performance.

Insert Table 8 around here

Table 8 summarizes participants' responses with regard to how they felt about the number of times they repeated the tasks. Between 28% and 53% of the participants indicated that six repetitions of the tasks were useful in improving their performance. The participants who reported that their performance did not continue to improve until the end were asked to specify the number of repetitions that they thought was sufficient in each case. As reflected in Table 8, the sub-set of respondents (47% on the narration, 59% on the opinion; 72% on the instruction) indicated that three or four performances were sufficient. This thus triangulates the results of the fluency analysis which showed improvement most clearly across the first four performances of each task. Furthermore, it indicates that participants generally did not become bored or fatigued, but appreciated the chance to complete the same tasks with different partners across the time frame of a typical lesson.

DISCUSSION AND CONCLUSIONS

In answer to the first research question concerning the effect of aural-oral task repetition with different interlocutors on short-term L2 fluency, our findings indicate that: (1) participants' speech rate improved markedly over the first three performances and then gradually until the fifth performance, (2) clause-final pausing decreased only between the first two performances, (3) mid-clause pausing decreased in a step-wise fashion between the first and third, second and fourth, and the third and fifth performances, and (4) significant changes in overt self-repair behavior were only detected between the fifth and first two performances and the sixth and first three performances respectively. Immediate massed repetition of input-based and output-based versions of tasks in a sequence may thus have differential effects on L2 speech production. These findings have implications for past and future research on task repetition as it relates to theories of L2 speech production mechanisms and attention during performance.

The findings suggest that up to five performances of a task may be required to prime, activate and optimize students' linguistic encoding processes on a task so that they can avoid breakdown and monitor their performance efficiently. Our results for speech rate provide additional evidence for the beneficial effects task repetition on enhancing fluency. They are in line with previous research that has found significant gains in speech rate when tasks were repeated once (Ahmadian & Tavakoli, 2011; Arevart & Nation, 1991, 1993; Bygate, 2001; Lynch & Maclean, 2000; Wang, 2014) and several times (Ahmadian, 2011). However, speech rate is likely to reflect efficiency in carrying out all aspects of L2 speech production: content planning, linguistic encoding, articulation, and monitoring (Götz, 2013; Kormos, 2006). Performance of the task for the first time builds a macro-structure conceptual plan for the content to be conveyed (Skehan, 2014). It also assists the speaker in the linguistic formulation of their message (Levelt, 1989, 1999) in terms of priming lemmas in the mental lexicon as well as the syntactic, morphological and phonological information associated with them, and activating syntactic building procedures (Skehan, 2014). Wang (2014), for example, has shown that even the articulation phase of speech production benefits from previous performance. These positive effects of previous performance on all three major processes of speech production are reflected in the large gains in speech rate in the present study. However, the other measures of fluency observed provide more nuanced insights into the benefits of repetition.

Our results for clause-final pausing complement Wang's (2014) findings in that the average length of pauses at clause boundaries decreased after one repetition of our task. Previous research has shown that clause-final pausing is used by speakers for content-planning, activating background knowledge and task-specific schemas, as well as for ordering the information to be conveyed (Butterworth, 1975; Götz, 2013). Hence, it is closely connected with the conceptualization stage of L2 speech production (Levelt, 1989, 1999; Kormos, 2006). The results for the present study indicate a reduction in clause-final pausing for all learners on the first task repetition suggesting that one of the benefits of the first task repetition is the enhancement of the conceptualization process.

The fact that no significant difference was detected in mid-clause pause frequency between the first and second performance in the present study is also parallel to the findings of Wang (2014) who found no reduction in the length of mid-clause pausing after only one repetition. Between the first and third performances, however, the participants in our study reduced mid-clause pausing. Mid-clause pausing has been found to signal breakdowns in the linguistic encoding process and tends to occur because of difficulties in lexical access or syntactic encoding (Götz, 2013; Kormos, 2006). In particular, mid-clause pausing reflects difficulties in retrieving relevant lemmas and accessing the morpho-syntactic information associated with them (Götz, 2013; Skehan et al., 2012). Decreased mid-clause pausing is thus an indicator of more efficient linguistic encoding mechanisms. These encoding mechanisms can be facilitated by priming effects from the student's own previous performances as well by the lexical items and syntactic constructions in their partner's speech (McDonough & Trofimovich, 2009; Pickering & Ferreira, 2008). The finding that mid-clause pausing diminished significantly only between the first and third performances and this reduction was only partially overlapping with changes in clause-final pausing suggests that dividing attentional resources efficiently between conceptualization and linguistic encoding processes is challenging for L2 learners.

With regard to the frequency of overt self-repairs, decreases were only observed in the fifth and sixth performances. It can be argued that the frequency of self-repairs is indicative of how much attention speakers devote to monitoring processes of L2 speech production (Kormos, 1999). As participants optimize their conceptualization and encoding processes, it could be argued that they need to pay less attention to monitoring. The results for the present study provide some support for this hypothesis. The reduced frequency of self-repairs in the fifth and sixth performances suggests that the participants' accuracy and efficiency in linguistic encoding had improved by this point so that fewer reformulations and overt self-corrections were necessary.

Our second and third research questions then asked whether any observed effects of aural-oral task repetition would vary across tasks of differing discourse demands and speakers of differing proficiency levels respectively. MANOVA indicated no significant interaction between task and repetition overall suggesting that this form of task repetition affected fluency similarly regardless of differing discourse and structural task demands. MANOVA also confirmed that aural-oral task repetition with different interlocutors had positive effects on speed, breakdown and repair fluency for participants at all three levels of proficiency, which demonstrates the robustness of task repetition on L2 performance. Even language learners at relatively high levels of proficiency (C1 and B2 levels) can be expected to improve their fluency and optimize the use of their linguistic resources and speech encoding processes as a result of repeating the same task as speakers and listeners with different interlocutors within a given instructional sequence. It is possible, however, that because the linguistic encoding processes of higher level learners are more automatized than those of lower level learners, the allocation of attentional resources and the benefits of task repetition for conceptual and linguistic processing vary at different levels of proficiency. Therefore, future research which investigates the linguistic production of students together with their pausing profile is needed to uncover potential proficiency-related benefits of task repetition pertaining to the division of attentional resources at different stages of speech production.

Our final research question then investigated participants' perceptions concerning the benefits of aural-oral task repetition and how many repetitions of each task they found useful. Very few participants provided overt indications of fatigue in repeating these tasks six times with different partners. Benefits with regard to improvements in fluency were mentioned by a large number of participants, and they also reported instances of learning from their peers and of existing knowledge representations being primed. For the proportion of participants who did not feel that six performances of a task were necessary, the optimal number of repetitions they perceived as being useful for the narration task was four, slightly less for the opinion task, and closer to three for the instruction task. The nature of the tasks themselves might partly explain these results. The narration task has a clear structure together with variation in content from frame-to-frame. Success is perceived when the story is told to conclusion, and the variation in the content allows learners to refine their resources in each stage of the story. In other words, the task is structured enough for learners to focus their efforts, but at the same time challenging enough to incorporate variation. The opinion task, on the other hand, is open-ended. As participants hear entirely different opinions about the photos each time they repeat the task, they may perceive a plateau in their improvement. Repetition may thus have been perceived to be less useful (see Table 8). Finally, the instruction task is much easier than the other two tasks due to the fact that the content is very repetitive across the picture frames, and the plateau effect might have been felt at an earlier stage. Participants' responses thus generally triangulate the results of the L2 production data.

Two primary limitations need to be kept in mind in interpreting the findings of the study. First, the decision to collect the data in a classroom context with students working in pairs with each other rather than being recorded performing the monologue tasks alone in a laboratory environment resulted in a level of background noise that made it impossible to analyze the location and duration of unfilled pauses reliably with automated software. The results of the study are thus limited in that only filled clause-final and mid-clause pauses were considered. Furthermore, controlled studies are now needed to reveal how additional features of fluency such as the mean length of runs and the length of unfilled pauses are affected by massed task repetition. Second, the decision to have learners work with each other and alternate roles of speaker and listener across the task sequence resulted in the observed effects for task repetition being combined with the effects of interlocutor input and any number of other interlocutor variables such as gender, personality, conversation style, etc. Although efforts were made to distribute these effects across the design through partial counterbalancing (see Procedures above), the study combines the effect of task repetition with that of interlocutor input. The decision to have students work in pairs and alternate speaker and listener roles might also have contributed to the lack of boredom and fatigue expressed in participants' responses to the tasks.

As we argued earlier in this paper, however, the benefit of the approach used in the study was that it preserves the integrity of the tasks as communicative tools. It also adds external or ecological validity to the results in that they are more reflective of task repetition as it is likely to occur inside and outside the classroom. Nevertheless, it is necessary that the results be interpreted in terms of optimizing speech production in the short term when input-based and output-based tasks are repeated with different interlocutors in a classroom situation as the order in which learners received input from speakers of different proficiency levels varied across the six repetitions of the task set, and this variation could have affected uptake at different points in the sequence. Future research on the effects of repetition independent of interlocutor input and other interlocutor effects is now needed to provide a more fine-grained picture of the effects of repetition as 'practice' rather than repetition as 'communication' on the efficiency of L2 speech production.

ACKNOWLEDGEMENTS

We are grateful for the support of the Research Fund of Lancaster University's Faculty of Arts and Social Sciences. This publication is supported by datasets, which for ethical and legal reasons cannot be publicly shared.

NOTES

1. Covert repairs, when errors are intercepted before articulation, are very difficult to identify reliably because they might not be directly observable in spontaneous speech (Kormos, 1999), and hence they were not investigated in the current study.

REFERENCES

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

Arevart, S. (1989). Grammatical change through repetition. *RELC Journal, 20,* 42-60.

Arevart, S. & Nation, P. (1991). Fluency improvement in a L2. *RELC Journal, 22,* 84-94.

Arevart, S. & Nation, P. (1993). Adjusting fluency and grammar through repetition. In Oller, J. (Ed.), *Methods that work: Ideas for language teachers* (2nd edition) (pp. 298-308). Boston: Heinle & Heinle.

Bosker, H., Pinget, A., Quené, H., Sanders, T., & De Jong, N. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing, 30*, 159–175.

Butterworth, B. (1975) Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research, 4,* 75-87.

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

Bygate, M. (1999). Task as context for the framing, reframing and unframing of language. *System, 27,* 33-48.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Council of Europe. (2001). *Common European Framework of reference of languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

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

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Linguistic skills and speaking fluency in a L2. *Applied Psycholinguistics, 33*, 1-24.

Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in L2 Acquisition, 20,* 349-385.

Educational Testing Service. (2008). TOEIC Test Data and Analysis 2007: Number of Examinees and Scores in FY2007. Tokyo: The Institute for International Business Corporation, TOEIC Steering Committee.

Ellis, N. C. (2002). Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in L2 Acquisition, 24,* 143–188.

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford: Oxford University Press.

Ellis, R. (2005). Planning and task-based research: Theory and research. In R. Ellis (Ed.), *Planning and task-performance in a L2* (pp. 3–34). Amsterdam: John Benjamins.

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics,* 19, 221-246.

Ellis, R. & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*, 354–75.

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

Götz, S. (2013). *Fluency in native and nonnative English speech* (Vol. 53). Philadelphia/Amsterdam: John Benjamins Publishing.

Kormos, J. (1999). Monitoring and self-repair in L2. *Language Learning, 49,* 303-342.

Kormos, J. (2006). Speech production and L2 acquisition. Mahwah, N.J.: Lawrence Erlbaum.

Kormos, J. (2011). Speech production and the Cognition Hypothesis. In P. Robinson (Ed.), *L2 task complexity: researching the cognition hypothesis of language learning and performance* (pp. 39-60). Amsterdam: John Benjamins.

Lennon, P. (2000). The lexical element in spoken L2 fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25–42). Ann Arbor, MI: University of Michigan Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1999). Producing spoken language: a blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), Neurocognition of language (pp. 83-122). Oxford, England: Oxford University Press.

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

Lynch, T. & Maclean, J. (2001). A case of exercising: effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan & M. Swain. (Eds.) *Researching pedagogic tasks* (pp. 141-162). Harlow: Longman.

Maurice, K. (1983). The fluency workshop. *TESOL Newsletter, 17,* 429.

McDonough, K., & Trofimovich, P. (2009). *Using priming methods in second language research*. London: Taylor & Francis.

Nihon Eigo Kentei Kyokai [The Eiken Foundation of Japan] (2009). *Eiken jun 1 kyuu zenmondaishuu* [*Eiken Pre-1 Level: Complete Questions Collection*]. Tokyo: Oubunsha.

Pang F., & Skehan P. (2014). Self-reported planning behaviour and L2 performance in narrative retelling. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 95-128). Amsterdam: John Benjamins.

Pickering M.J., & Ferreira V.S. (2008). Structural priming: A critical review. *Psychological Bulletin, 134,* 427-459.

Plonsky, L., & Oswald, F.L. (2014). How big is big? Interpreting effect sizes in L2 research. *Language Learning* 64, 878-912.

Segalowitz, N. (2010). *Cognitive bases of L2 fluency*. New York: Routledge.

Skehan, P. (2009). Modelling L2 performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics, 30*, 510–32.

Skehan, P. (2014). Limited attentional capacity, second language performance and task-based pedagogy. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 211-260). Amsterdam: John Benjamins.

Skehan, P., Xiaoyue, B., Quian, L., & Wang, Z. (2012). The task is not enough: Processing approaches to task-based performance. *Language Teaching Research, 16,* 170-187.

Tabachnick, B. & Fidell, S. (2007). Using multivariate statistics 5th ed. Boston: Pearson

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, *61*, 37-72.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a L2* (pp. 239-277). Amsterdam: John Benjamins.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

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: John Benjamins.

Wigglesworth, J. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14*, 85–106.

Yuan, F. & Ellis, R. (2003). The effect of pre-task planning and online planning on fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, *24*, 1–27.

Yule, G. (1997). *Referential communication tasks*. Hillsdale, NJ: Lawrence Erlbaum.

Appendix. Sample of pausing and overt self-repair analyses

…because they have a trouble with **er** (*FP mid*) a graffiti problem. And **it's just it is not just** (*REP*) at this time. They had this problem before. So **uhm** (*FPbound*) then, they tried to solve this problem. So they have a meeting with **er** (*FP mid*) shop owners. But they think about the solutions. But the security cameras and street patrol is good way to solve it, but **it's it needs** (*REP*) too much cost. So they can't get nice idea at this meeting. But that night, a woman **go went** (*REP*) back her home. And when she **find found** (*REP*) **she's her** (*REP*) son **is was** (*REP*) writing some picture, suddenly she come up with nice idea which **er** (*FP mid*) she think graffiti is problem, but if it's art, **it doesn't it's not** (*REP*) problem. **er** *(FPbound*) so they gather children a month later **er** *(FPbound*) and make children write some picture on her shop's wall before **er** (*FP mid*) graffiti problem comes again.

FP mid – Filled pause in mid-clause position

FP bound – Filled pause at clause boundary

REP- overt self-repair

Table 1.

Descriptive Statistics for TOEIC Scores across Groups

TOEIC Scores	N	Range	Min.	Max.	Mean	SD	Skewness (z-score)	Kurtosis (z-score)
Band A (860-990)	8	85	885	970	916.88	29.147	.850	.028
Band B (730-860)	11	125	730	855	795.91	43.579	-.767	-.898
Band C (470-730)	13	245	470	715	591.92	65.241	.097	.212
Combined	32	500	470	970	743.28	144.287	-.522	-1.518

Table 2

Pairing and Ordering of Participants for Each Performance of the Task Set

Performances	Pairing and ordering of speakers			
1st Time	A-B	D-C	2-1	4-3
2nd Time	C-A	B-D	3-1	2-4
3rd Time	A-D	C-B	4-1	2-3
4th Time	1-A	B-4	3-C	D-2
5th Time	A-2	1-B	C-4	3-D
6th Time	3-A	B-2	1-C	D-4

Table 3

Speech Rate Means and Standard Deviations across Tasks, Performances and Proficiency Levels (n=32)

		Instruction		Narration		Opinion	
Performance	Proficiency	M	SD	M	SD	M	SD
1	High	2.19	.60	2.10	.46	1.64	.26
	Mid	1.58	.37	1.65	.38	1.28	.33
	Low	1.29	.40	1.40	.28	1.07	.26
	Total	1.62	.56	1.66	.45	1.29	.36
2	High	2.33	.55	2.20	.42	1.97	.39
	Mid	1.90	.53	1.79	.31	1.58	.46
	Low	1.54	.40	1.51	.23	1.40	.47
	Total	1.86	.57	1.78	.41	1.60	.49
3	High	2.80	.68	2.31	.27	2.20	.37
	Mid	1.94	.40	2.06	.30	1.90	.44
	Low	1.69	.40	1.85	.31	1.59	.38
	Total	2.05	.65	2.04	.34	1.85	.46
4	High	2.87	.63	2.48	.49	2.22	.26
	Mid	2.13	.40	2.17	.39	1.81	.35
	Low	1.82	.34	1.92	.27	1.60	.33
	Total	2.19	.60	2.14	.42	1.83	.39
5	High	2.89	.62	2.75	.65	2.47	.43
	Mid	2.21	.41	2.30	.41	1.97	.29
	Low	1.88	.42	1.96	.35	1.73	.37
	Total	2.24	.61	2.27	.54	2.00	.46
6	High	2.87	.68	2.65	.34	2.40	.34
	Mid	2.24	.47	2.31	.35	2.03	.32
	Low	1.92	.29	2.05	.29	1.65	.34
	Total	2.27	.59	2.29	.39	1.97	.44

Table 4.

Clause-final pausing means and standard deviations across tasks, performances and proficiency levels (n=32)

		Instruction		Narration		Opinion	
Performance	Proficiency	M	SD	M	SD	M	SD
1	High	.0095	.0115	.0170	.0204	.0179	.0184
	Mid	.0097	.0099	.0143	.0102	.0185	.0160
	Low	.0126	.0192	.0238	.0205	.0296	.0213
	Total	.0108	.0144	.0188	.0176	.0228	.0191
2	High	.0086	.0066	.0104	.0084	.0115	.0104
	Mid	.0081	.0112	.0120	.0095	.0219	.0157
	Low	.0075	.0075	.0125	.0144	.0245	.0201
	Total	.0080	.0085	.0118	.0112	.0204	.0170
3	High	.0029	.0041	.0096	.0062	.0082	.0066
	Mid	.0102	.0094	.0085	.0068	.0106	.0115
	Low	.0104	.0106	.0134	.0107	.0253	.0268
	Total	.0085	.0093	.0108	.0085	.0159	.0198
4	High	.0036	.0084	.0109	.0088	.0052	.0049
	Mid	.0070	.0083	.0053	.0071	.0116	.0100
	Low	.0101	.0105	.0137	.0122	.0159	.0139
	Total	.0074	.0093	.0101	.0103	.0117	.0114
5	High	.0043	.0061	.0054	.0051	.0083	.0050
	Mid	.0067	.0079	.0039	.0047	.0104	.0105
	Low	.0110	.0142	.0156	.0126	.0166	.0123
	Total	.0079	.0107	.0090	.0103	.0124	.0106
6	High	.0027	.0037	.0070	.0058	.0081	.0047
	Mid	.0055	.0054	.0079	.0078	.0119	.0076
	Low	.0078	.0079	.0155	.0145	.0165	.0147
	Total	.0057	.0064	.0107	.0111	.0128	.0109

Table 5

Mid-clause pausing means and standard deviations across tasks, performances and proficiency levels (n=32)

		Instruction		Narration		Opinion	
Performance	Proficiency	M	SD	M	SD	M	SD
1	High	.0066	.0089	.0094	.0067	.0145	.0127
	Mid	.0113	.0123	.0177	.0161	.0171	.0107
	Low	.0263	.0224	.0288	.0212	.0373	.0303
	Total	.0162	.0183	.0201	.0181	.0247	.0233
2	High	.0061	.0058	.0151	.0109	.0124	.0139
	Mid	.0107	.0102	.0148	.0105	.0105	.0108
	Low	.0152	.0174	.0228	.0169	.0184	.0132
	Total	.0114	.0131	.0181	.0137	.0142	.0127
3	High	.0063	.0090	.0098	.0100	.0072	.0081
	Mid	.0081	.0072	.0114	.0090	.0114	.0085
	Low	.0133	.0134	.0188	.0169	.0135	.0118
	Total	.0098	.0106	.0140	.0133	.0112	.0099
4	High	.0042	.0083	.0091	.0078	.0099	.0111
	Mid	.0050	.0101	.0094	.0107	.0154	.0131
	Low	.0152	.0148	.0137	.0138	.0108	.0112
	Total	.0089	.0127	.0111	.0114	.0122	.0117
5	High	.0067	.0095	.0061	.0060	.0097	.0113
	Mid	.0046	.0051	.0069	.0080	.0052	.0074
	Low	.0124	.0139	.0167	.0150	.0152	.0122
	Total	.0083	.0107	.0107	.0120	.0104	.0111
6	High	.0050	.0075	.0062	.0037	.0114	.0092
	Mid	.0028	.0041	.0069	.0078	.0061	.0096
	Low	.0119	.0080	.0123	.0097	.0189	.0194
	Total	.0070	.0078	.0089	.0082	.0126	.0150

Table 6

Self-repair means and standard deviations across tasks, performances and proficiency levels (n=32)

		Instruction		Narration		Opinion	
Performance	Proficiency	M	SD	M	SD	M	SD
1	High	.0144	.0106	.0140	.0081	.0196	.0162
	Mid	.0192	.0179	.0222	.0154	.0194	.0178
	Low	.0262	.0198	.0268	.0157	.0360	.0207
	Total	.0208	.0174	.0220	.0146	.0262	.0199
2	High	.0144	.0094	.0158	.0137	.0158	.0108
	Mid	.0189	.0110	.0201	.0148	.0167	.0099
	Low	.0235	.0109	.0321	.0131	.0237	.0131
	Total	.0196	.0109	.0239	.0152	.0193	.0117
3	High	.0112	.0136	.0138	.0136	.0147	.0089
	Mid	.0224	.0157	.0212	.0150	.0161	.0093
	Low	.0212	.0136	.0264	.0175	.0334	.0197
	Total	.0191	.0146	.0214	.0161	.0228	.0166
4	High	.0144	.0121	.0156	.0074	.0223	.0134
	Mid	.0087	.0089	.0113	.0074	.0202	.0114
	Low	.0217	.0169	.0213	.0130	.0291	.0203
	Total	.0154	.0142	.0164	.0107	.0244	.0161
5	High	.0137	.0138	.0099	.0091	.0123	.0129
	Mid	.0098	.0090	.0130	.0102	.0162	.0105
	Low	.0229	.0177	.0252	.0156	.0283	.0145
	Total	.0161	.0150	.0172	.0139	.0201	.0143
6	High	.0122	.0116	.0137	.0094	.0142	.0060
	Mid	.0111	.0085	.0155	.0064	.0143	.0100
	Low	.0152	.0140	.0216	.0141	.0233	.0114
	Total	.0131	.0115	.0176	.0110	.0180	.0105

Table 7.

Perceptions of task repetition

Task Type	Usefulness in general	N	Usefulness for fluency	Usefulness for incorporation	Other	N
Opinion	100% (28)	28	41% (12)	48% (14)	10% (3)	29
Narration	96.8% (30)	31	42% (13)	35% (11)	23% (7)	31
Instruction	74.2% (23)	31	40% (10)	24% (6)	36% (9)	25

Table 8

Perceptions of Number of Productive Repetitions

Task Type	Improved until end (N=31)	Optimum Repetitions (Means)	N
Narration	53% (17)	4.0	14
Opinion	41% (13)	3.6	18
Instruction	28% (9)	3.2	22

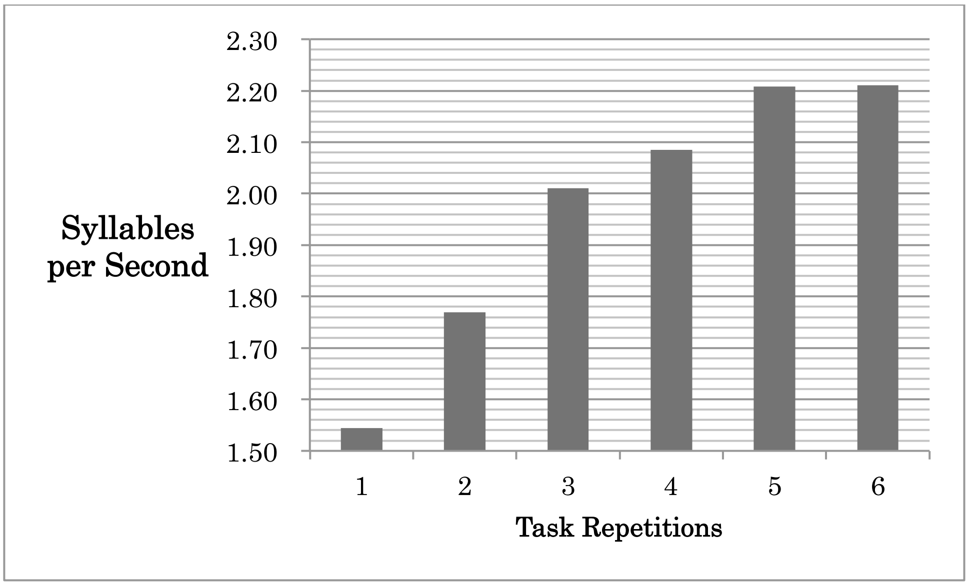

Figure 1 Repetition and speech rate

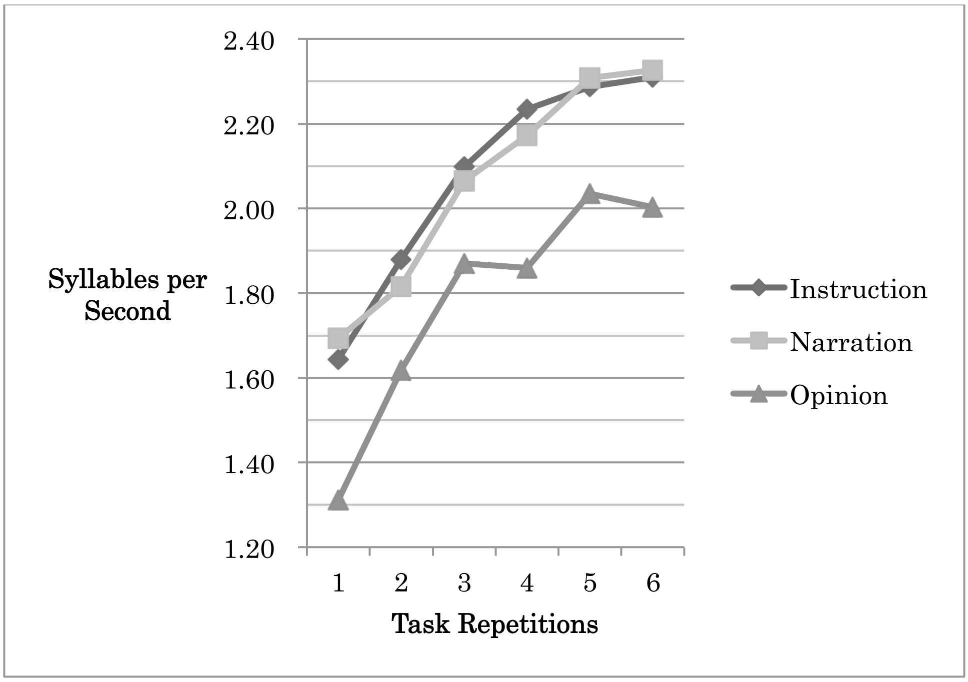

Figure 2 Repetition, task and speech rate

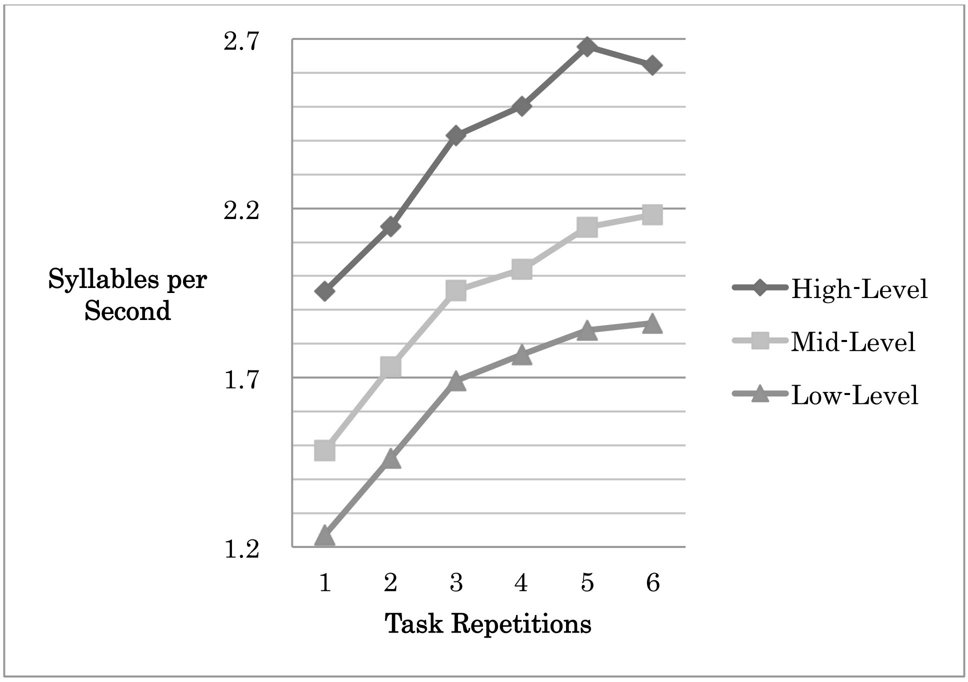

Figure 3 Repetition, proficiency and speech rate

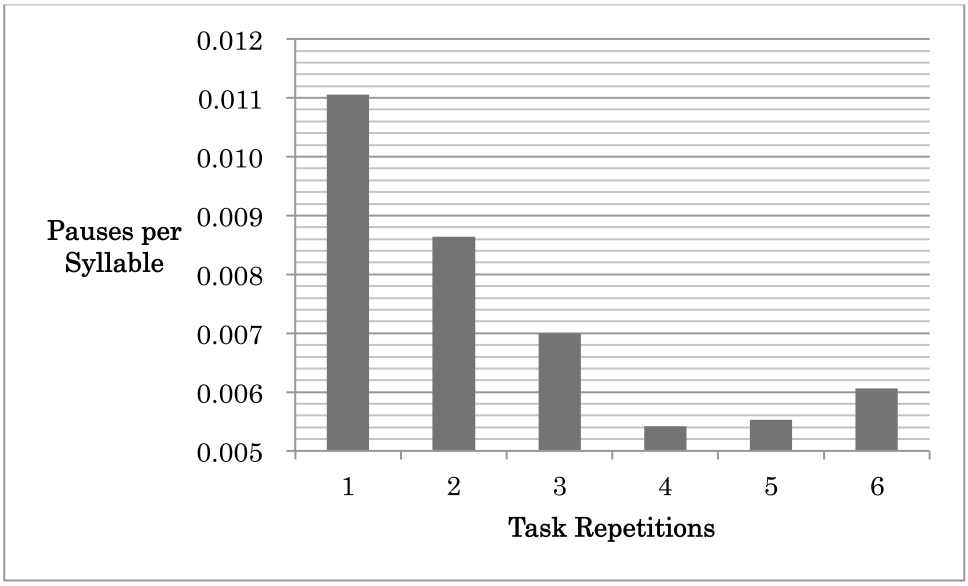

Figure 4 Repetition and clause-final pausing

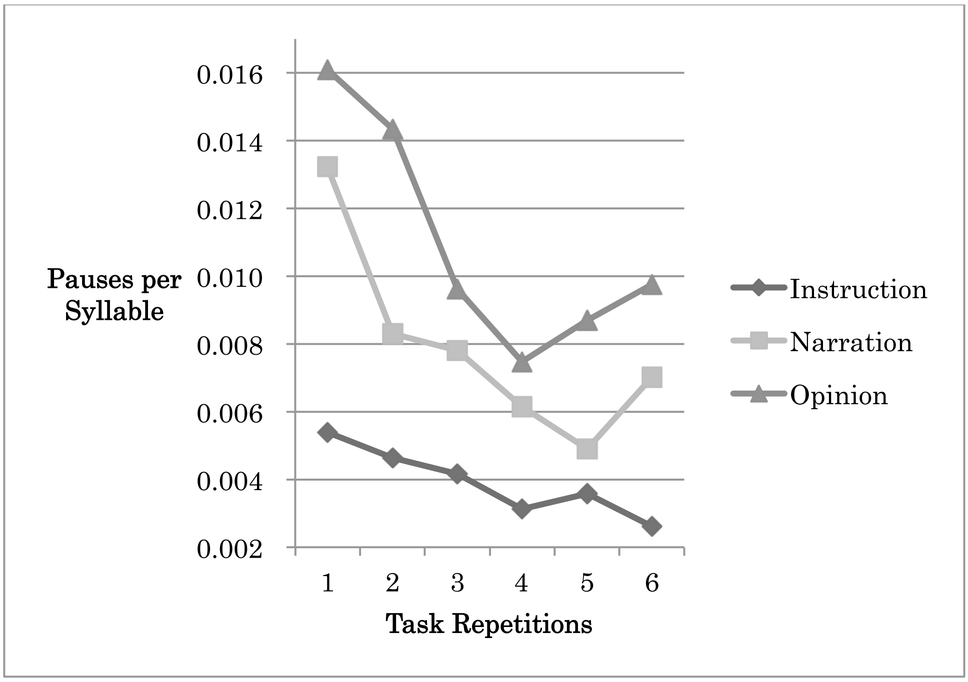

Figure 5 Repetition, task and clause-final pausing

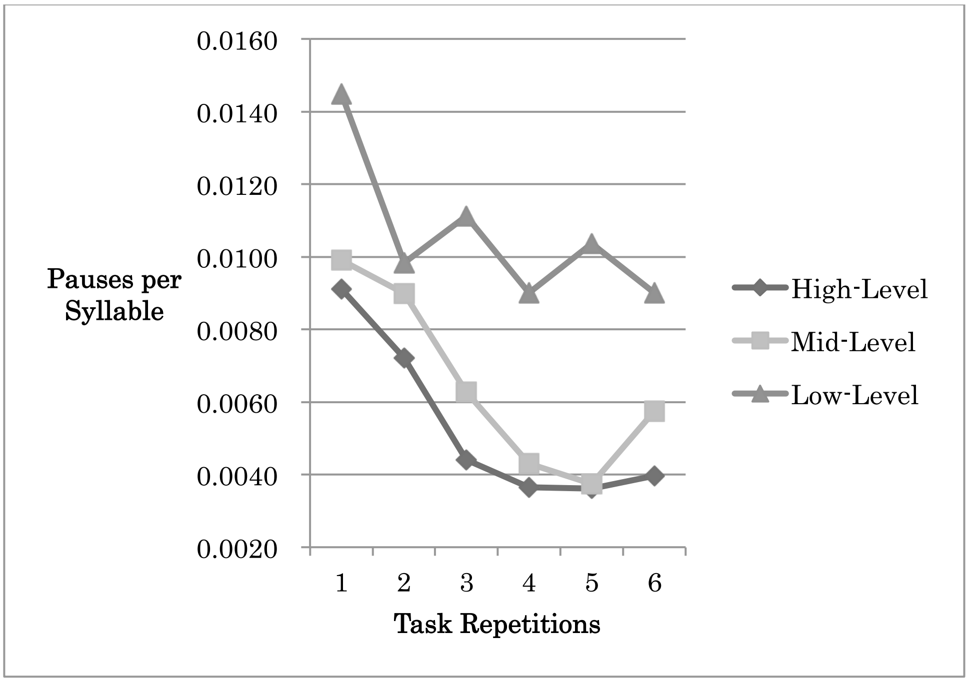

Figure 6 Repetition, proficiency and clause-final pausing

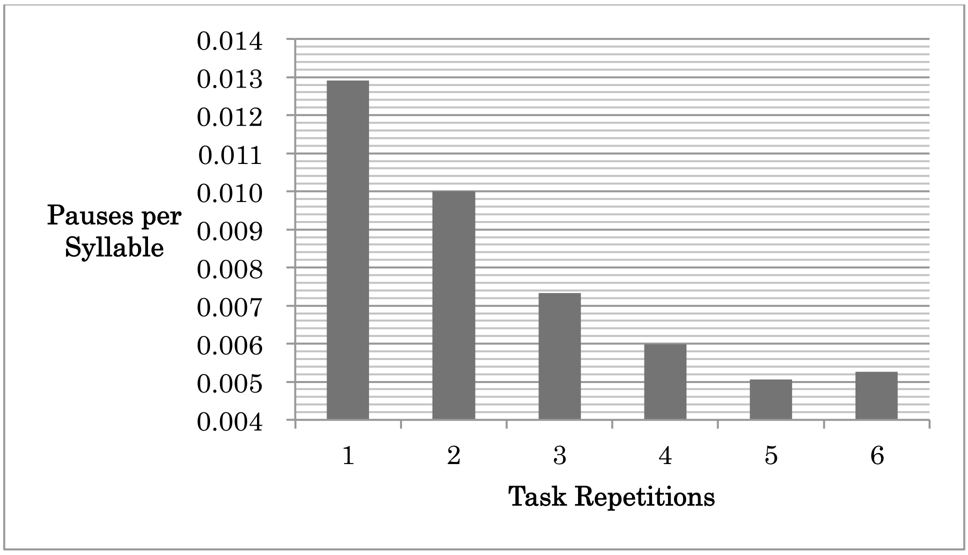

Figure 7 Repetition and mid-clause pausing

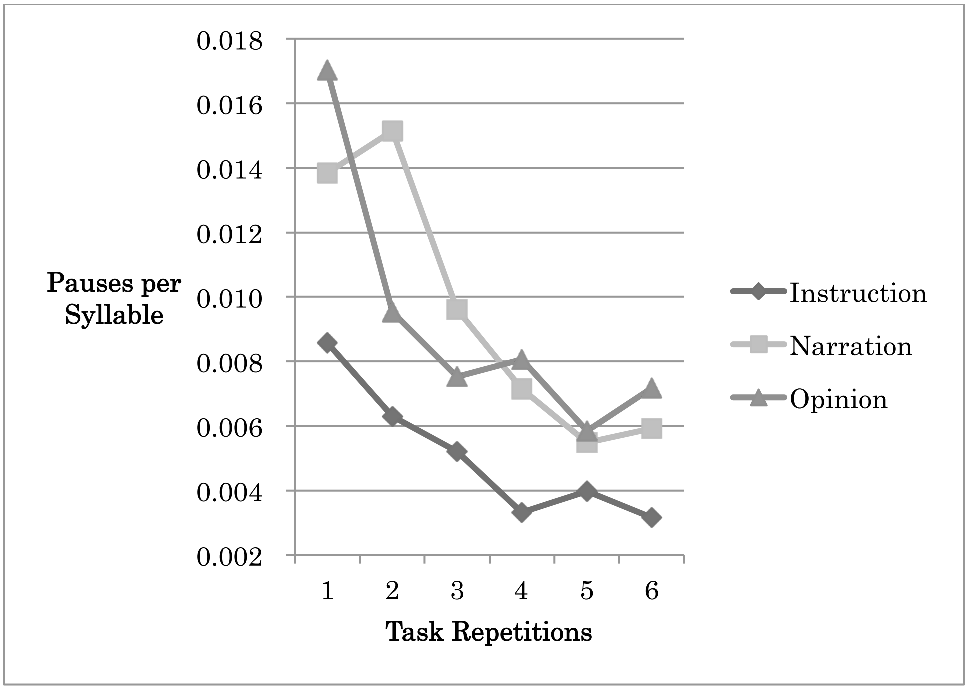

Figure 8 Repetitions, task and mid-clause pausing

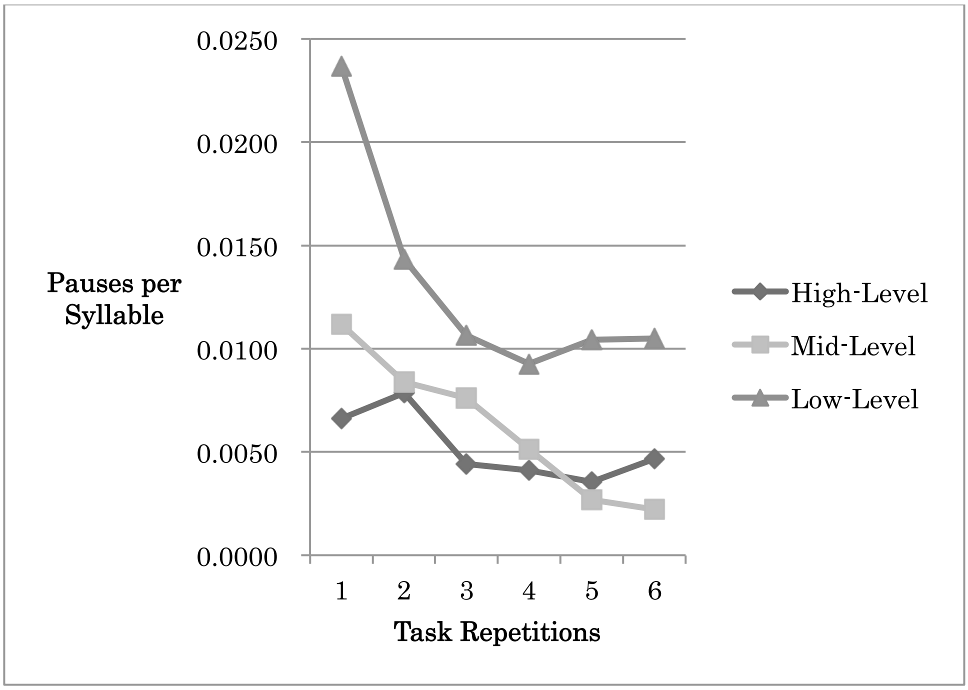

Figure 9 Repetitions, proficiency and mid-clause pausing

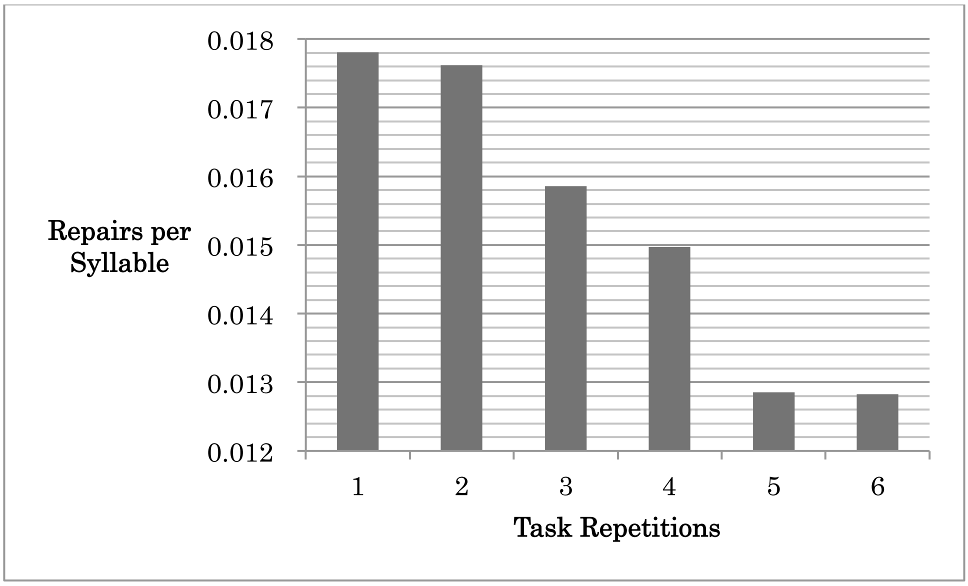

Figure 10 Repetition and self-repair

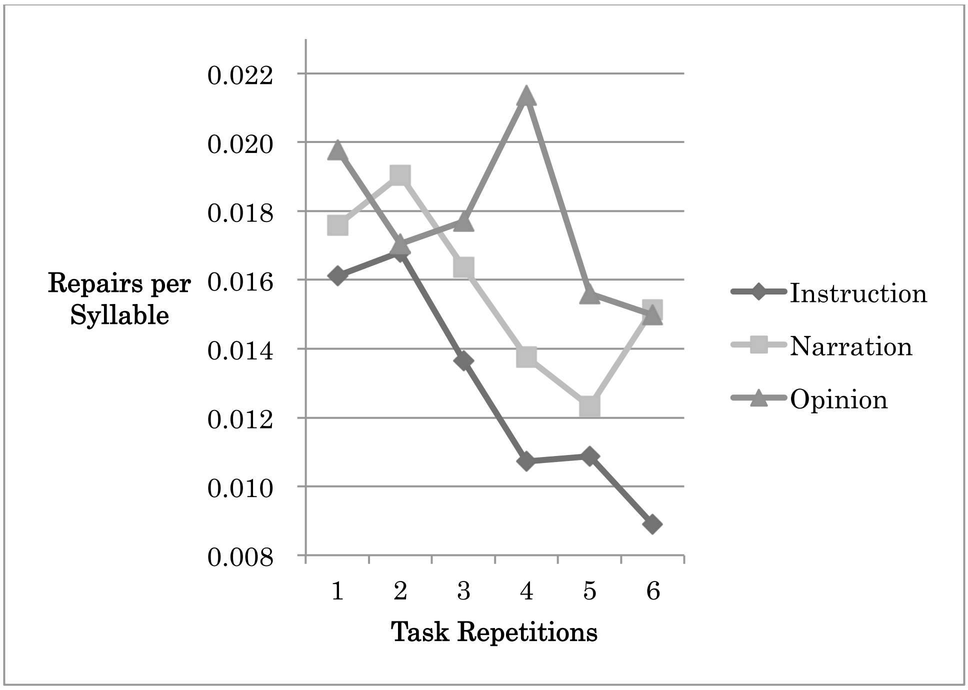

Figure 11 Repetitions, task and self-repair

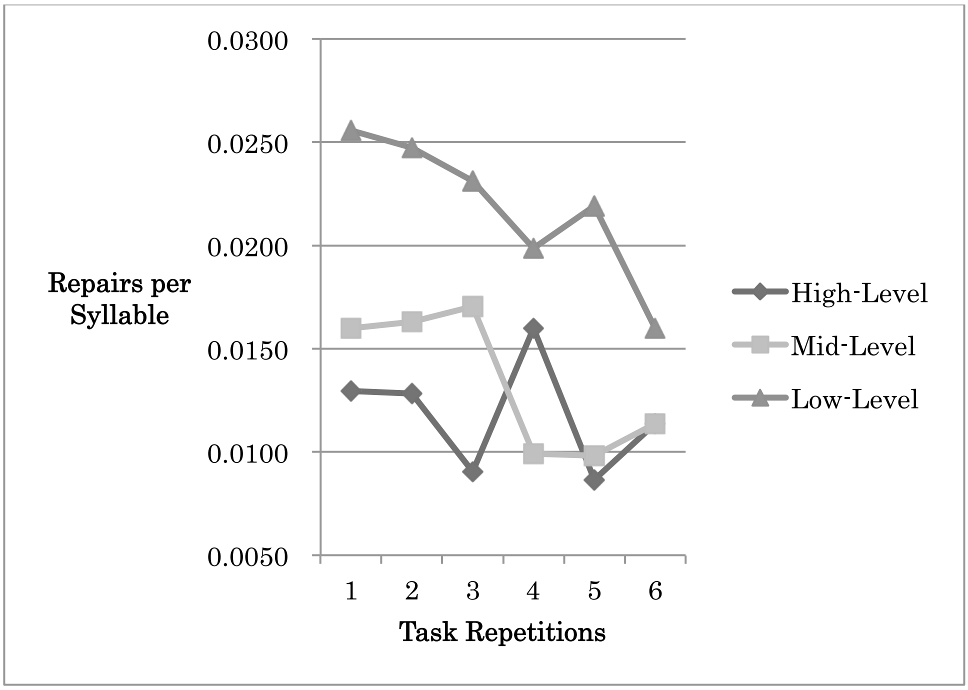

Figure 12 Repetitions, proficiency and self-repair