LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

Fluency Training in the ESL Classroom: An Experimental Study of Fluency Development and Proceduralization

Nel de Jong

Vrije Universiteit Amsterdam

Charles A. Perfetti

University of Pittsburgh

The present study investigates the role of speech repetition in oral fluency development. Twenty-four students enrolled in English-as-a-second-language classes performed three training sessions in which they recorded three speeches, of 4, 3, and 2 min, respectively. Some students spoke about the same topic three times, whereas others spoke about three different topics. It was found that fluency improved for both groups during training but was maintained on posttests only by the students who repeated their speeches. These students had used more words repeatedly across speeches, most of which were not specifically related to the topic. It is argued that proceduralization of linguistic knowledge represented a change in underlying cognitive mechanisms, resulting in improvements in observable fluency.

Keywords classroom research; English as a second language; fluency training; oral fluency; proceduralization; second language; task repetition; vocabulary

The ultimate goal of many second-language (L2) learners is to be fluent in the target language—that is, to be able to express their thoughts easily, with more

The data of the pretests and posttests were presented at the AAAL conference in Costa Mesa, CA, on April 24, 2007. The authors would like to thank Colleen Davy, Jessica Hogan, Rhonda McClain, and Laura Halderman for their contributions to transcription, coding, and analysis of the data from the training sessions. Thanks to Laura Halderman, Mary Lou Vercellotti, and Scott Walters for their feedback to drafts of this article. Funding for this research was provided by the National Science Foundation, grant No. SBE-0354420 to the Pittsburgh Science of Learning Center (PSLC; http://www.learnlab.org).

Correspondence concerning this article should be addressed to Nel de Jong, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands. Internet: cam.de.jong@let.vu.nl

DOI: 10.1111/j.1467-9922.2010.00620.x

attention to meaning than form, in any given situation. Communication should eventually be smooth, with some processes of production relatively fast and automatic. Although there is not a single agreed-upon definition in the literature, fluency is often understood to refer to the flow and smoothness of delivery (Chambers, 1997; Koponen & Riggenbach, 2000). Beyond this core idea are some distinctions about fluency that are sometimes made. Lennon (1990), for example, distinguished between broad and narrow fluency. In the broad sense, fluency is like general proficiency and includes accuracy and complexity of the output. In the narrow sense, however, fluency is restricted to temporal measures, such as length and number of pauses and the number of hesitations (e.g., *I uh like sports*) and repetitions (e.g., *I like I like sports*). Another distinction is made by Segalowitz (2000), who differentiated between cognitive fluency and performance fluency. The former concerns "the efficiency of the operation of the cognitive mechanisms underlying performance," whereas the latter refers to "the observable speech, fluidity, and accuracy of the original performance" (p. 202). Although performance fluency is the goal of many language learners, it is highly dependent on the knowledge and skills of the speaker, which are the bases of cognitive fluency. More specifically, fluency is related to the extent of a speaker's linguistic knowledge as well as the use of that knowledge, the speed of access, and control over the available linguistic forms and syntactic devices. However, because speed of access and control of linguistic forms become fluent only with much practice in speech production, how to support gains in this narrow sense of fluency within the reduced opportunities for practice provided by language classrooms is an important question and is the focus of the present research.

In the following sections, we will first review research on oral fluency in L2 learning settings. Next, we will give an overview of explanations of fluency and fluency development as well as measures of fluency.

Oral Fluency in the Context of L2 Learning and Teaching

Fluency and fluency development have been studied in a number of different contexts, including learning contexts, planning, task repetition, and language instruction. These include studies of immersion, study abroad, and language instruction in the home country (e.g., Freed, 1995; Freed, Segalowitz, & Dewey, 2004; Segalowitz & Freed, 2004; Towell, 2002; Towell, Hawkin, & Bazegui, 1996). Emphasizing the limitations of classroom opportunities, the results of such studies favor immersion settings with home-country language courses producing the least progress in the measures of rate of speech and length of runs. DeKeyser (2007) concluded that a "majority of [study-abroad] students

make measurable progress in speaking, especially in terms of fluency, at least in the programs of longer duration" (p. 211). However, he also argued that many students make less progress than expected, which is often due in part to the limited quantity and quality of opportunities for proceduralization and automatization of rule use.

Gains in oral performance, however, can be achieved through procedures that theoretically affect the processes of fluent production. Both pretask planning (Foster & Skehan, 1996; Mehnert, 1998; Yuan & Ellis, 2003) and task repetition (Bygate, 2001; Bygate & Samuda, 2005; Lynch & Maclean, 2000, 2001) support fluency by freeing up attentional resources. Task repetition, for example, supports speakers' selection of words, morphemes, and grammatical structures. Bygate (2001) found that repetition affected performance even when speeches were 10 weeks apart. Complexity was higher in the repeated task, but there was a trade-off with fluency, which may have been due to a shift of attention to complexity. Importantly, fluency was measured only as the number of unfilled pauses per t-unit; results may have been different if different measures had been used. Finally, no beneficial effect was found on a repeated task that had a different topic. Immediate repetition, on the other hand, was shown by Lynch and Maclean (2000, 2001) to improve accuracy in terms of phonology, vocabulary, semantic precision, and syntax. In addition, there was some evidence for gains in fluency.

Despite the research on fluency in task-based learning, few studies have investigated how instructional techniques affect the mechanisms underlying the longer term development of fluency. A notable exception is the study by Snellings, Van Gelderen, and De Glopper (2002). They showed that classroom instruction can increase lexical retrieval speed—a process of language production that supports fluency—even with only 10 encounters over a period of 4 weeks. A follow-up study provided evidence for transfer to a narrative writing task, in that more of the trained words were used (Snellings, Van Gelderen, & De Glopper, 2004). However, no effect on global quality was found, and writing fluency was not examined. Gatbonton and Segalowitz (2005) argued that automaticity can be promoted by inherently repetitive tasks that elicit formulaic language and are genuinely communicative and thus fit well in a Communicative Language Teaching classroom setting, although they report no data that support this conclusion.

The instructional technique that is the focus of the present study was specifically designed for oral fluency development: the fluency workshop, or 4/3/2 procedure (Arevart & Nation, 1991; Maurice, 1983; Nation, 1989; Wood, 2001). Although this technique has been used in classrooms for several decades, few

studies have empirically investigated its effects so far. In this task, students speak about a given topic for 4 min and then retell it twice, as close to verbatim as possible, in 3 and 2 min, respectively. The 4/3/2 involves task time pressure and repetition, but in contrast to the studies by Bygate and others discussed earlier, the speeches are repeated immediately. In that way, the speakers have the additional benefit of having used certain vocabulary and grammatical constructions, which can facilitate retrieval through lexical and syntactic priming (e.g., Bock & Loebell, 1990; Branigan, Pickering, & Cleland, 2000; McDonough & Mackey, 2006; Pickering & Branigan, 1999; Youjin & McDonough, 2008).

Nation (1989) investigated the fluency, accuracy, and complexity of speeches given in the 4/3/2 task, comparing the first and last speeches. He found an increase in speech rate (words per minute) and a decrease in the number of false starts, repeated words, and hesitations (such as *uh, um*). Accuracy improved only slightly for half of the participants, mostly when grammatical contexts were repeated but not for errors that involved inflections. The strategies used by the speakers to fit their speeches into less time included omitting unimportant details and changing grammatical constructions, which in some cases involved more complex sentences. Arevart and Nation (1991) replicated the study with a greater number of participants and found that both speaking rate (words per minute) and hesitations per minute improved significantly on the retellings. They concluded that the 4/3/2 task gives learners the opportunity to speak with higher than normal fluency and complexity during their third delivery. Neither study tried to tease apart the effects of repetition and time pressure, nor did they include posttests to examine the long-term effects of the task, in contrast to the present study.

In summary, the study of fluency has mostly focused on short-term effects instead of longer term development. Whereas short-term effects on fluency can be explained by planning and repetition, which enable the speaker to shift attention and to benefit from priming, longer term effects may require proceduralization and automatization, which will be discussed in more detail in the following section.

An Information Processing Perspective on Oral Fluency

Fluency development is often explained in terms of procedural knowledge and automatic processes, as when Schmitt (1992) characterized fluency as "an automatic procedural skill" (p. 358) and Segalowitz (2000) argued that observable fluency reflects the balance between automatic and controlled processes (p. 214). Procedural knowledge, or "knowledge how," is different from

declarative knowledge, or "knowledge that." Declarative knowledge includes not only knowledge of word forms and encyclopedic knowledge but also of explicit grammatical or phonological rules, and it is generally slower to use and requires more attention and cognitive resources than procedural knowledge. Because procedural knowledge is processed fast and in parallel with other processes and because it puts less of a burden on the limited resources of working memory, it is more suitable for fluent speech. The declarative/procedural distinction is well known in cognitive psychology and neuroscience (e.g., Anderson, 1983; Anderson et al., 2004; Anderson & Lebiere, 1998; Squire, 1987, 1992) and has been applied to L2 acquisition (e.g., De Jong, 2005; DeKeyser, 1997; Ferman, Olshtain, Schechtman, & Karni, 2009; Hilton, 2008; Towell, 2002; Towell et al., 1996; Ullman, 2001a, 2001b, 2004). It has been argued that, in fluent L1 and L2 speakers, procedural knowledge is mostly involved in the encoding stages of language production (e.g., phrase and clause structure building) and in articulation, whereas declarative knowledge is associated with retrieval of lexical items and their syntactic information as well as conceptualizing and monitoring (Kormos, 2006; Levelt, 1989, 1999; Towell et al., 1996). Less advanced L2 learners may still rely on declarative knowledge, which they need to proceduralize.

The description of the process of proceduralization is found in the ACT-R model (Anderson & Lebiere, 1998). According to this model, declarative knowledge takes the form of chunks, which are small independent patterns of information (e.g., 3 + 4 = 7, or the English regular past tense morpheme is "–ed"). These chunks can stem from encoding an object in the environment, like remembering a word, or they can be the result of a previously executed production (e.g., after computing that 3 + 4 equals 7, the information "3 + 4 = 7" is encoded as a chunk; after creating *work* + –ed, *worked* is encoded as a chunk). Procedural knowledge consists of *production rules*, which correspond to steps of cognition and have the basic form of "goal condition + chunk retrieval => goal transformation." For instance, if the goal is to add 3 and 4, the chunk "3 + 4 = 7" can be retrieved, and then the goal can be changed to the next goal (e.g., the next step in a multicolumn addition). In ACT-R, each production rule is triggered by a goal and retrieves one or, at most, a few declarative chunks. Production rules are the units of skill acquisition and each has its own learning curve. Declarative chunks and production rules are competing with other chunks or rules, and the strongest one is triggered. Strength is determined, among other things, by amount and recency of use.

Anderson et al. (2004) argued that the change in retrieval speed of declarative chunks follows a power law, with initial practice leading to large gains and

later practice leading to gains that gradually diminish. Eventually, performance moves toward an asymptote. Proceduralization involves the creation of production rules and combining smaller production rules into larger ones. These new production rules subsequently need to gain strength so as to be able to compete with other, previously existing, production rules. Strength can be gained by repeated practice.

In the context of language, we could say that the retrieval speed of words and phrases increases with repeated practice, with large initial gains that gradually diminish to smaller gains. The creation and strengthening of new chunks can lead to the emergence of formulaic sequences. Language use can also lead to the construction of new production rules and the collapsing of production rules into larger ones. Repeated practice is necessary for these collapsed production rules to be able to compete with (and defeat) their "parent" production rules.

Measures of Fluency

The combination of several measures, as used in the present research, can give evidence of chunking and proceduralization, as explained below. First, there is the mean length of pauses measured in seconds. The different ways of determining pauses and setting cutoff points are discussed below. Second, the phonation/time ratio is calculated as the percentage of time spent speaking as a proportion of the total time taken to produce the speech sample. This measure is related to the number of pauses in a speech: If the mean length of pauses is stable but the number of pauses decreases, phonation/time ratio increases. Third, the mean length of fluent runs is the mean number of syllables produced between pauses. Finally, the articulation rate—in syllables per minute—is calculated by dividing the total number of syllables produced by the amount of time taken to produce them, excluding pause time. It is slightly different from speech rate, which includes pause time. Kormos and Dénes (2004) found that the first three of these measures were good predictors of fluency ratings by native and nonnative speaker judges, although articulation rate was not. (Two other measures not included in this study were also good predictors: speech rate and pace—that is, the number of stressed words per minute.)

Towell et al. (1996) argued that these measures in combination can be used as indicators of proceduralization. The number and length of pauses by themselves are not reliable indicators of proceduralization, as they vary with task demands, planning opportunities, and speaker characteristics (some speakers pause more and longer than others). Another measure to consider is the mean length of fluent runs (i.e., stretches of speech that are spoken without pauses).

In theory, when proceduralization has taken place, learners are able to produce longer fluent runs, but a speaker can also produce longer runs by taking more time for planning, which may show up as longer pauses. Therefore, if the mean length of fluent runs increases while the mean pause length and phonation/time ratio are stable, more silent planning time was not needed, which indicates that encoding and sentence building have been proceduralized.[1] Hence, mean length of fluent runs may be used as an indicator of proceduralization when it is used in combination with the mean length of pauses and phonation/time ratio. Finally, articulation rate is a measure of the speed of articulatory processes and is thus not strongly related to the proceduralization of lexical and syntactic knowledge.

Purpose of the Study

Despite the evidence for improvements due to task repetition discussed earlier, it is not obvious whether the 4/3/2 procedure would result in a long term effect and transfer to new topics. Short-term effects can be explained by reduced cognitive load due to priming. However, long-term effects and transfer to new topics would only be expected if the processes underlying speech production were changed as a result of the 4/3/2 task (e.g., through proceduralization of knowledge). Accordingly, the goal of this study is to examine the effectiveness of the 4/3/2 task with a focus on long-term effects, transfer, and proceduralization.

The 4/3/2 task has two main features: time pressure and speech repetition. Time pressure may encourage students to express their ideas more quickly and efficiently, with shorter pauses and more complex language structures. Speech repetition, which is our focus here, may lead to an increase in fluency because of advantages at several levels. First, at the semantic level (conceptualization), students generate the content for their delivery during their planning time and while they speak. In the second and third deliveries, they can benefit from this, which removes the need to pause and hesitate to plan new semantic content. Second, vocabulary and grammatical structures are generated not so much during pretask planning as during the first delivery of a speech (cf. Bygate & Samuda, 2005, p. 65). Again, in subsequent deliveries students can benefit from having generated content. Even though they may not be able to remember and reproduce their first delivery verbatim, the words and grammar items they used are still more activated than before and thus more readily available for use (the priming effect discussed earlier). Pauses related to lexical searches and hesitations related to monitoring of grammar are likely to be reduced for these

items. Overall, when students repeat their speech, they do not have to generate content (semantic, grammatical, lexical), which frees up cognitive resources, which can be used in several ways. One way is to speak more fluently, with shorter pauses and fewer hesitations, as Nation (1989) found. Another way is to get access to different language items, such as more sophisticated, or specific, vocabulary and more complex grammatical structures, which is also consistent with Nation's findings.

Beyond affecting the accessibility of specific content (declarative knowledge), speech repetition may support proceduralization. For example, if a student uses a grammatical item, like a relative clause or an embedded question, which he/she knows but does not use very often, it might take some time to use it. Monitoring and corrections may be needed to arrive at the correct construction (e.g., *I don't know what is his name uh what his name is*). However, if the learner can encode the new chunk and reuse the item in subsequent speeches, this may strengthen its representation and accelerate retrieval. A new production rule may be formed to retrieve the chunk. Similarly, a word that may not be available when there is a heavy burden on working memory during the first speech may become available during a subsequent delivery, when more cognitive resources are available. For example, a student may first say that shopping *takes a long time* and in a repeated delivery be more specific in saying that it *wastes time*. Using the word *waste* will strengthen its representation and make it more available on subsequent occasions.

The present study, in order to focus on the effect of speech repetition, compared students who repeated their speeches with students who spoke about different topics each time. Speech repetition was expected to increase cognitive fluency during the 4/3/2 procedure, because it would—temporarily—increase the availability of vocabulary and sentence structures, which would lead to shorter pauses and a higher phonation/time ratio. This would leave more cognitive resources for other processes, which, in turn, would lead to a longer mean length of fluent runs. In addition, it was expected that cognitive fluency would increase long term, due to the repeated practice. Because the longer term effect of the training was measured in new speeches about different topics, any indication of proceduralization must be ascribed to something broader than changes in the processing of topic-related vocabulary only. If no indication of proceduralization were to be found and instead only speed of articulation were to increase, this should show up as an increase in articulation rate only.

The overall research question was whether the 4/3/2 task would lead to a long-term increase in fluency through proceduralization, leading to the following hypotheses.

1. Gains in fluency during the 4/3/2 task and from pretest to posttest would show evidence of proceduralization, more so for students who repeated their speeches than for those who did not.
2. Any gains in fluency would persist over 1–4 weeks and would transfer to new topics, more so for students who repeated their speeches than for those who did not.

The sources of the fluency gains were examined by studying the fluency measures and vocabulary use across speeches in the 4/3/2 task.

In contrast to Nation's (1989) study, in the present study students worked individually on a computer. By not having students work in pairs, the task was less naturalistic, but because the influence of a conversation partner was eliminated, control over the training task was increased. An additional benefit for the students and teachers was that there was more time for each student to speak, as they only took on the role of speakers, not of listeners.

Method

Participants

This study took place in an institute for English as a second language (ESL) at a large university in the United States during the fall semester of 2006.[2] All 47 students enrolled in Speaking courses at a high intermediate level (level 4, three sections) agreed to participate. Students are typically placed at this level when they have a score of 60–79 on the Michigan Test of English Language Proficiency (MTELP; Corrigan et al., 1979). Placement is also based on an in-house listening test and a writing sample. Most students are simultaneously enrolled in reading, writing, listening, and grammar courses at the same institute. Due to absences, only 24 complete datasets were available for analysis. Four students chose not to disclose their background information, including gender, age, and first language. Of these remaining 20 students, 7 were female and 13 were male. The age range was 19–37, with an average of 25 years. All tasks—pretests and posttests as well as all training sessions—were part of the students' regular class requirements. First languages spoken included Arabic (8), Chinese (5), French (2) Korean (2), and single speakers of Spanish, Japanese, and Portuguese.

Materials

In the 4/3/2 task, students spoke about a given topic for 4 min, then for 3 min, and, finally, for 2 min. The pretest and posttests were part of the regular course curriculum and consisted of 2-min recorded monologues. These tests

were part of a larger graded activity for which students also transcribed their speeches and commented on their accuracy, and teachers gave feedback. These transcriptions, comments, and feedback were not part of the present study. All topics, in both the training and the test sessions, were of general interest to the student population at the language institute and included topics such as "What do you think about pets?" and "Who is your favorite artist?" The topics were followed by a few additional questions in order to give students more suggestions for the contents of their speech (see the Appendix). There may have been variability in the difficulty of the topics, but this had minimal impact on the analyses that focused on between-subjects comparisons. Moreover, the topic of the last 4/3/2 fluency training session was identical to the pretest topic (Pets) in order to track progress on comparable topics.

Procedures

The participants were randomly assigned to one of three conditions. The first condition was the Repetition condition: students performed the original 4/3/2 task, in which they spoke about one topic three times. In the second condition (No Repetition), students performed the same task as the Repetition condition but spoke about three different topics. The third condition was the Repetition-II condition, in which students performed exactly the same tasks with the same topics as in the Repetition condition but later in the semester. The Repetition-II condition therefore served as a control condition (no training) for the first part of the study (comparing Test 1 and Test 2), whereas in the second part of the study (between Test 2 and Test 3), this was a Repetition condition. Note that Test 2 was the immediate posttest for the Repetition and No Repetition conditions, whereas it was a pretest for the Repetition-II condition. Students of two sections were randomly assigned to the Repetition and No Repetition conditions. Students in a third section were all assigned to the Repetition-II condition. All students performed the 4/3/2 task three times over a period of 2 weeks.

The pretests and posttests were 2-min speeches given 2 or 3 days before the first training session (pretest), as well as approximately 1 and 4 weeks after the last training session (immediate and delayed posttests). All training and test sessions took place during regular class hours. The tests and fluency training each had their own introductory session to familiarize the students with the procedures. Table 1 presents the schedule for the tests and training sessions.

Fluency Training Sessions (4/3/2)

A common misconception among the students, according to the course administrators, was that fluency is nothing more than speaking fast. In order to avoid

Table 1 Schedule of tests and training sessions

	Test	Training sessions			Test	Training sessions			Test
	1	A	B	C	2	A	B	C	3
Repetition	a	1, 1, 1	4, 4, 4	7, 7, 7	b				c
No Repetition	a	1, 2, 3	4, 5, 6	7, 8, 9	b				c
Repetition-II	a				b	1, 1, 1	4, 4, 4	7, 7, 7	c

Note. The numbers and letters refer to the topics as listed in the Appendix.

students trying to speak as fast as possible, which may have prevented other effects of repetition, students were given a short description of fluency at the beginning of each 4/3/2 training session. The description included reference to the temporal factors of length and number of pauses as well as planning and the use of familiar vocabulary and grammar.

In each 4/3/2 training session, students were given a topic and instructed to "make a few notes about what you want to say" and "don't write sentences, but only a few keywords." Time for note-taking was 3–5 min, after which students were encouraged to continue. This pretask planning time was provided for students to generate enough semantic content to fill 4 min and was given for each new topic (not for repeated topics). Audio recordings of the students' speeches were made with the software (16 bits, 22 Hz, one-channel sound quality). During the speeches, the topic and the student's notes were presented on the screen, but the notes could no longer be edited. A clock indicated the elapsed time.

After the first delivery, students were asked a number of evaluation questions, which were designed to have them reflect on their performance. Students checked the boxes of statements about what they would do differently next time, pertaining to the factors mentioned in the description of fluency (e.g., "I should use only words that I know well" and "I should pause fewer times"). Other open-answer questions concerned general performance, such as which difficult words the students wanted to remember and what superfluous information they would not include in the next speech. Next, students spoke again, this time for only 3 min. Students in the Repetition and Repetition-II conditions were asked to repeat their speech, whereas students in the No Repetition condition saw a new topic and took new notes. After this second delivery, evaluation questions followed that asked students to compare their performance to the first delivery. Then another delivery followed, this time for 2 min only. The session ended with evaluation questions and a brief questionnaire about the students' general

performance in the session. In this questionnaire, students gave ratings on a 7-point scale (1 = disagree; 7 = agree) to 13 statements, including "I knew what words I wanted to say" and "This exercise was useful."

All training sessions were run by the first author, and the class teacher assisted with keeping the students on task. Because of the classroom setting, and at the request of the teachers, feedback was provided to the students after each session. Individual reports were written by the first author and handed out in the next session. These reports included the student's average length of pauses, the number of short pauses (<1 s) and long pauses (>1 s), as well as the overall averages of the students in the study.

Test Sessions

The test sessions were part of the regular curriculum and, therefore, run by the regular teachers and graduate student assistants. The topics for the pretests and posttests were selected from among the three topics discussed in the week before the test. Students were not informed which topic would be on the test. In contrast, none of the topics in the fluency training were discussed beforehand.

The pretests and posttests were run on Apple PowerMac computers, and the training sessions were run on Dell personal computers. The software for the tests and training was developed with the Revolution Studio 2.6.1. package (Shafer, 2006).

Analyses

All speeches were first transcribed with pauses indicated so that the mean pause length and the phonation/time ratio could be calculated. Syllables were counted to compute the mean length of fluent runs. The transcriptions were coded for parts of speech and retracings (e.g., repetitions, corrections), so that the number of word types repeated across deliveries could be calculated. Analyses were performed on the data of students who completed all training and test sessions: 10 students in the Repetition condition, 9 students in the No Repetition condition, and 5 students in the Repetition-II condition. Because of the small group size of Repetition-II, this group was collapsed with the Repetition condition in the analyses of the training data.

Transcription and Pauses

All speeches from the pretests and posttests were transcribed by the first author, using PRAAT 4.6.06 (Boersma, 2001). The speeches from the 4/3/2 training sessions were transcribed by two trained research assistants who checked each other's work. To find pauses, the beginning and end of each speech segment

was determined first by using the PRAAT function "To textgrid (silences)." All pause boundaries were checked and adjusted by the transcribers as necessary, by listening to the recording and visually inspecting the spectrogram and waveform. Nonverbal fillers such as "uh," "ah," "um," and "mmm" were transcribed and treated as pauses.

A pause was defined as silence or a nonverbal filler of 200 ms or longer. This cutoff point is slightly lower than the 250–400 ms that other researchers use (e.g., Freed et al., 2004; Goldman-Eisler, 1961; Segalowitz & Freed, 2004; Towell et al., 1996), but it follows Lennon (1990) because the majority of the pauses of 200 ms and longer sounded dysfluent. Although the pause length and phonation/time ratio computed in this study include both fluent and dysfluent pauses, any decrease in pause length or increase in the phonation/time ratio is likely to be due mainly to changes in the number and length of dysfluent pauses. The upper limit to pauses was set to 2.5 standard deviations above the mean in a student's particular speech, which is a fairly conservative and commonly used criterion for eliminating outliers. Pauses that were longer than this were replaced by the mean plus 2.5 standard deviations. The trimming was necessary because some students may have been briefly distracted; very infrequently, students needed to be encouraged to continue speaking. Such long pauses, usually around 3 or 4 s, would not be an indication of the students' fluency or proceduralization. After the pauses were determined, the phonation/time ratio was computed by dividing the total time filled with speech (not including silent pauses and nonverbal fillers like "uh") by the total time spent speaking (time filled with speech + pauses and nonverbal fillers).

Syllable Counting

In order to calculate the length of fluent runs, syllables were counted by a research assistant. Where there was doubt about the number of syllables pronounced (e.g., "every" can be pronounced as /ɛvri/ or /ɛvəri/), the original recording was consulted. False starts were counted as syllables, but fillers such as "uh," "um," and "mmm" were not. To obtain a reliability measure, syllable counts of 36 min of speech selected randomly from the 4/3/2 training data were recounted by the first author. The percentage agreement between the two counts was 96%. Where there were discrepancies, the difference was usually only one syllable.

Retracings and Parts of Speech Coding

Prior to the lexical analyses, retracings were coded and part-of-speech tags were added to the transcripts, using the CLAN software (http://childes.psy.cmu.edu/).

Retraced words and syllables were included in the syllable counts but not in the vocabulary analysis. Three types of retracings were coded: without correction (e.g., *it's [/] um it's [/] it's like a dog*), with correction (e.g., *<the fish is> [//] the fish are swimming*), and with reformulation (e.g., *all of my friends had [///] uh we had decided to go home for lunch*).[3] The MOR and POST programs of the CLAN software were run to generate part-of-speech tags for each word. These tags were needed to generate accurate vocabulary lists for the lexical analysis (e.g., to distinguish between the noun and verb *travel*). Ambiguities were resolved by a trained research assistant who had also transcribed the speeches, and the tags were checked by a second assistant.

Statistical Analyses

General linear model (GLM) analyses with repeated measures were used to analyze the fluency data of the 4/3/2 speeches and the pretests and posttests. Separate GLMs were performed for each measure: mean length of fluent runs, phonation/time ratio, mean length of pauses, and articulation rate. For the pretest and posttest data, the within-subjects variable was time (Test 1, 2, and 3) and the between-subjects variable was condition (Repetition, No Repetition, and Repetition-II). For the analyses of the data from the training sessions, multivariate GLMs were used with the three sessions (Session A, B, and C) as measures. Planned post hoc univariate GLMs were performed to analyze the effects within the training sessions. The within-subjects variable was delivery (Delivery 1, 2, and 3; respectively, the 4-min, 3-min, and 2-min speeches). The two Repetition conditions (Repetition and Repetition-II) were collapsed because they involved the same training tasks. The between-subjects variable for the training tasks, therefore, was condition (Repetition, No Repetition). Univariate analyses for each session are also reported. The alpha level for all statistical tests was set at .05.

Lexical Analysis

In order to assess the extent to which the students repeated their speeches—at least in terms of vocabulary—the amount of overlap in vocabulary between pairs of speeches ("lexical overlap") was calculated by computing the number of words that were used in all three speech deliveries within a session, in two deliveries, or in only one. Only lexical words (nouns, verbs, adjectives, and adverbs) were included in this analysis; retracings were not included. In addition, correlations were computed between the number of repeated words and gain scores of the four temporal measures of fluency from pretest to immediate posttest. Finally, the number of repeated topic-related and topic-unrelated words was compared.

Results

We first present the results of the pretests and posttests to assess evidence for proceduralization (hypothesis 1) and for long-term retention and transfer to speeches about different topics (hypothesis 2). Next, three additional analyses are presented to examine the source of the fluency gains.

Pretest/Posttest Data

To test the proceduralization hypothesis 1, the temporal measures were analyzed to find evidence for longer fluent runs with stable or improved length of pauses and phonation/time ratios. In addition, articulation rate was examined to see if any improvement concerned speed only. To test the long-term retention and transfer hypothesis 2, we examined whether gains were retained over 4 weeks and transferred to different topics.

Proceduralization

Table 2 presents for each of the three conditions the three measures of proceduralization: mean length of fluent runs (in syllables), mean length of pauses (in seconds), and phonation/time ratio. Note that the students in the No Repetition and Repetition conditions performed the 4/3/2 training between Test 1 and

Table 2 Means and standard deviations of the measures of proceduralization in the pretests and posttests

	Test 1	Test 2	Test 3
Mean length of fluent runs (in syllables)			
Repetition[a]	4.54 (1.34)	4.83 (1.79)	4.70 (1.53)
No Repetition[b]	4.42 (1.55)	4.12 (1.01)	4.28 (.80)
Repetition-II[c]	4.58 (1.25)	4.80 (.95)	5.44 (1.20)
Mean length of pauses (in seconds)			
Repetition	1.12 (.29)	0.94 (.21)	0.97 (.25)
No Repetition	.87 (.13)	1.00 (.27)	0.92 (.11)
Repetition-II	.94 (.07)	0.92 (.11)	0.84 (.11)
Phonation/time ratio			
Repetition	54.75 (12.13)	60.22 (10.51)	58.63 (9.25)
No Repetition	59.34 (8.85)	55.50 (8.82)	57.11 (7.78)
Repetition-II	57.00 (4.24)	58.63 (5.51)	62.43 (7.67)

[a] $n = 10$.
[b] $n = 9$.
[c] $n = 5$.

Test 2, whereas the Repetition-II condition performed the training between Test 2 and Test 3.

The GLM analyses showed a significant interaction between time and condition for pause length, $F(4, 42) = 3.897$, $p = .009$, partial $\eta^2 = .271$, and phonation/time ratio, $F(4, 42) = 2.563$, $p = .052$, partial $\eta^2 = .196$, but not for mean length of fluent runs. A series of post hoc two-tailed *t*-tests was performed, comparing Test 1 and Test 2, and Test 2 and Test 3, for each measure in each condition. The Repetition condition showed significant differences only between Test 1 and Test 2 for mean pause length and phonation/time ratio, $t(9) = 3.647$, $p = .005$; $t(9) = 2.932$, $p = .017$, respectively. Although the small number of students in the Repetition-II condition was a concern, the difference in mean length of fluent runs between Test 2 and Test 3 still reached significance, $t(4) = 3.189$, $p = .033$, whereas the difference between Test 1 and Test 2 did not. No significant differences were found for the No Repetition condition. These results showed that in the Repetition condition, mean pause length decreased and the phonation/time ratio increased, whereas for the Repetition-II condition, mean length of fluent runs increased and pause length and the phonation/time ratio were stable. In both conditions, performance only changed over the time interval in which the fluency training had taken place. As will be discussed in more detail later, we take both of these patterns of results as evidence for proceduralization and thus support for hypothesis 1. Performance in the No Repetition condition did not change over time.

Articulation Rate

Articulation rate, presented in Table 3, was measured as the number of syllables per minute of speech (pauses excluded). It is considered a measure of speed, unrelated to proceduralization. The main effect of time was significant, $F(2, 42) = 7.232$, p = .002, partial $\eta^2 = .256$, but the effect of condition was not.

Table 3 Means and standard deviations of the articulation rate (in syllables per minute) in the pretests and posttests

	Test 1	Test 2	Test 3
Repetition[a]	194.34 (25.21)	195.49 (33.35)	204.22 (40.64)
No Repetition[b]	193.91 (22.92)	189.82 (22.84)	203.53 (18.89)
Repetition-II[c]	206.94 (27.68)	214.44 (35.04)	232.69 (24.03)

[a] $n = 10$.
[b] $n = 9$.
[c] $n = 5$.

There was no interaction between time and condition. Post hoc two-tailed t-tests comparing Test 1 and Test 2, and Test 2 and Test 3 did not reveal any significant results, except the comparison between Test 2 and Test 3 under the No Repetition condition, which showed a trend, $t(8) = 2.255$, $p = .054$, indicating an increase in articulation rate. However, this increase occurred after training had ended and may thus be due to the students' regular language classes. The fluency training therefore did not seem to effect an increase in speed.

To summarize the students' performance on the tests, the Repetition-II condition shows the pattern of performance as expected, with an increase in length of fluent runs in combination with a stable mean length of fluent runs and stable phonation/time ratios. This is an indication of proceduralization, which enables learners to produce longer fluent stretches of speech without additional time for pausing. The Repetition condition shows a similar pattern in that length of fluent runs is stable while mean pause length and phonation/time ratios improve. This indicates that students were producing the same length of fluent stretches of speech but needed less pause time. Again, this could be considered evidence for proceduralization, contrasting with the results of the No Repetition condition, which showed no change in the proceduralization measures. The evidence thus supports both the proceduralization hypothesis 1 and the retention and transfer hypothesis 2, with gains observed over 4 weeks and for new topics (see the Discussion section for a fuller discussion of the evidence for both hypotheses).

Training Data

To examine whether gains would be observed within deliveries during training, the same fluency measures were applied to the training sessions. Multivariate repeated-measures GLM analyses were performed for each of the measures of proceduralization. Each session was a separate measure (training sessions A, B, and C). Delivery (the 4, 3, and 2-min deliveries) was a within-subjects independent variable and condition (Repetition vs. No Repetition) was a between-subjects variable. Because the Repetition ($n = 10$) and Repetition-II ($n = 5$) conditions performed exactly the same training—only at a different time—and because they both showed indications of proceduralization in their pretest/posttest data, their training data were collapsed ($n = 15$) for these analyses. However, for comparison with the test data, Tables 4 and 5 show the measures for the Repetition and Repetition-II condition separately.

Proceduralization

Table 4 shows the three measures of proceduralization per condition. For *mean length of fluent runs*, the multivariate analysis revealed a significant main effect

Table 4 Means and standard deviations of the measures of proceduralization during the 4/3/2 training sessions

	Session A Delivery			Session B Delivery			Session C Delivery		
	1	2	3	1	2	3	1	2	3
Mean length of fluent runs (in syllables)									
Repetition[a]	4.28	4.32	4.98	4.59	5.27	5.30	4.03	4.34	4.70
	(1.26)	(1.15)	(1.73)	(1.25)	(1.77)	(1.51)	(1.01)	(1.08)	(1.44)
No Repetition[b]	4.00	5.30	4.60	4.50	4.25	4.77	3.52	3.94	4.58
	(0.74)	(1.83)	(1.30)	(0.95)	(0.75)	(1.25)	(0.55)	(0.71)	(1.03)
Repetition-II[c]	4.62	4.73	5.55	5.16	5.37	5.51	5.02	5.12	5.42
	(0.83)	(0.92)	(1.52)	(0.96)	(0.78)	(1.13)	(1.35)	(0.85)	(1.72)
Mean length of pauses (in seconds)									
Repetition	1.10	1.08	0.91	0.96	0.98	0.95	0.99	0.89	0.82
	(0.28)	(0.27)	(0.21)	(0.21)	(0.31)	(0.19)	(0.23)	(0.20)	(0.22)
No Repetition	0.92	0.79	0.81	0.87	0.85	0.89	0.91	0.90	0.79
	(0.23)	(0.17)	(0.14)	(0.11)	(0.12)	(0.15)	(0.12)	(0.16)	(0.16)
Repetition-II	.99	0.84	0.78	0.83	0.78	0.82	0.88	0.81	0.76
	(0.22)	(0.15)	(0.07)	(0.12)	(0.07)	(0.10)	(0.22)	(0.20)	(0.10)
Phonation/time ratio									
Repetition	52.62	52.48	59.91	54.49	58.04	58.58	54.48	59.39	61.71
	(9.64)	(10.09)	(8.48)	(9.24)	(12.03)	(8.87)	(8.83)	(9.23)	(11.68)
No Repetition	55.07	62.23	60.52	59.04	58.32	59.55	53.12	55.05	61.13
	(10.30)	(10.42)	(8.44)	(6.26)	(6.68)	(7.79)	(7.58)	(7.11)	(8.26)
Repetition-II	53.86	58.99	62.39	60.69	63.10	62.68	59.13	62.81	63.44
	(7.52)	(5.10)	(5.41)	(6.19)	(5.50)	(3.43)	(8.70)	(8.17)	(6.81)

[a] $n = 10$.
[b] $n = 9$.
[c] $n = 5$.

of delivery and a significant interaction between delivery and condition, $F(6, 86) = 6.039$, $p = .000$, partial $\eta^2 = .296$; $F(6, 86) = 3.628$, $p = .003$, partial $\eta^2 = .202$, respectively. The univariate tests show that the main effects and interactions reached significance for all three sessions, except the interaction of delivery and condition in Session C. There seems to have been a fairly steady increase in length of fluent runs for the two Repetition conditions, whereas performance in the No Repetition condition was more variable, which is illustrated by the values in Table 4. For *mean length of pauses*, the multivariate analysis revealed only a significant effect of delivery, $F(6, 86) = 4.329$, $p = .001$, partial $\eta^2 = .232$, showing a decrease in pause length for all conditions. Univariate analyses showed the effect was only found in Sessions A and C, as the mean length of pauses decreased. For the *phonation/time ratio*, the multivariate analysis showed a significant main effect of delivery and a significant interaction between delivery and condition, $F(6, 86) = 6.516$, $p = .000$, partial $\eta^2 = .313$; $F(6, 86) = 2.633$, $p = .022$, partial $\eta^2 = .155$, respectively. The univariate tests showed that the main effect of delivery only reached significance for Session A and Session C, as the phonation/time ratio increased, and the interaction between delivery and condition reached significance only in Session A. Again, as illustrated by the data in Table 4, there appears to be a steady improvement for the two Repetition conditions (Repetition and Repetition-II) and more variable performance in the No Repetition condition.

Overall, the data from the training sessions reveal that the performance of all groups improved on all three measures, but mostly in the first and last training sessions (A and C). This may be due to increasing time pressure, as the available time decreased from 4 to 2 min. In addition, improvements were steadier for the two Repetition conditions than for the No Repetition condition, whose performance was more variable, probably because of the changes in topics during the sessions. In sum, the Repetition groups showed longer fluent runs with improved length of pauses and phonation/time ratios, but not over and above the improvements found in the No Repetition group. Interestingly, the questionnaire data showed that the majority of the students in the Repetition groups felt the second and third delivery were "easier" than the first, whereas fewer students did so in the No Repetition group. Nevertheless, the two conditions' ratings for the usefulness of the activity were very similar.

Articulation Rate

Table 5 shows the articulation rate during the 4/3/2 training session. The multivariate analysis revealed a significant main effect of delivery and a significant interaction between delivery and condition, $F(6, 86) = 6.549$, $p = .000$, partial

Table 5 Means and standard deviations of articulation rate (in syllables per minute) during the 4/3/2 training sessions

	Session A Delivery			Session B Delivery			Session C Delivery		
	1	2	3	1	2	3	1	2	3
Repetition[a]	202.45	205.08	211.00	214.32	220.11	225.69	196.04	196.26	205.30
	(27.96)	(29.67)	(32.09)	(29.66)	(32.81)	(30.24)	(27.31)	(29.19)	(27.45)
No Repetition[b]	203.35	223.78	213.33	205.48	206.89	209.35	197.69	208.71	216.85
	(14.92)	(20.69)	(16.77)	(13.68)	(15.40)	(19.87)	(23.13)	(25.04)	(24.29)
Repetition-II[c]	231.44	230.13	243.78	226.08	233.93	235.24	226.81	221.04	238.56
	(27.47)	(20.98)	(30.76)	(19.30)	(22.89)	(30.04)	(29.31)	(24.77)	(32.57)

[a] $n = 10$.
[b] $n = 9$.
[c] $n = 5$.

$\eta^2 = .324$; $F(6, 86) = 1.943$, $p = .035$, partial $\eta^2 = .156$, respectively. Univariate analyses showed that the effect of delivery and the interaction of delivery and condition were significant in all three sessions except the interaction in Session B. As Table 5 illustrates, there was an overall increase of articulation rate. Performance in the No Repetition condition was more variable, which suggests that articulation rate also depends on the topic of the speech, perhaps due to interest or familiarity.

Lexical Overlap

Having established the general effects of speech repetition on oral fluency development, we turn to a preliminary look at the possible role of word repetition in these effects. We calculated the number of words that were in only one, two, or all three speeches of a session. The results from Session A are presented in Table 6. It is clear that the students in the two Repetition conditions repeated many more word types across all three deliveries than students in the No Repetition condition, who used more word types in only one speech. However, the No Repetition group used a wider range of word types than the two Repetition groups, as their total number of word types was higher. The number of word tokens (not included in Table 6), on the other hand, was similar for the Repetition and No Repetition groups (162 and 170, respectively), whereas it was slightly higher for the Repetition-II group (190). However, a one-way ANOVA contrasting the three groups did not reveal any significant differences. In sum, more topic-related and topic-unrelated word types were used in all three deliveries under the two Repetition conditions than under the No Repetition condition.

In order to examine whether repeated use of vocabulary affected fluency gains, correlations were computed with the number of repeated word types and simple gain scores from pretest to immediate posttest. These correlations were computed for each of the four fluency measures (mean length of fluent runs, phonation/time ratio, mean length of pauses, and articulation rate). Table 7 shows there were moderate but significant correlations between the number of words used in three deliveries of a training session and pretest to posttest gains in the phonation/time ratio. Negative correlations were obtained with mean length of pauses. In Session C, correlations with these two fluency measures were also significant for words used in only two out of three deliveries. Correlations in Session B for phonation/time ratio with the number words used in three deliveries and in one delivery just missed significance but show a trend in the same direction as found in Sessions A and C.

Table 6 Means and standard deviations of words types repeated in the three deliveries of Session A

	Types used in three deliveries		Types used in two deliveries		Types used in one delivery	Total types
	All	Topic-related	All	Topic-related	All	All
A. Repetition[a]	23.9 (7.2)	7.7 (4.1)	21.7 (5.7)	3.8 (2.0)	47.2 (10.2)	93
B. No Repetition[b]	5.3 (2.2)	0.0 (0.0)	18.8 (4.1)	2.0 (2.3)	116.8 (18.6)	141
C. Repetition-II[c]	30.8 (3.1)	11.0 (2.1)	22.0 (9.7)	5.2 (3.8)	53.6 (12.1)	106
One-way ANOVA						
$F(2, 23)$	50.365	29.476	.672	2.688	62.662	13.984
p	.000	.000	.521	.091	.000	.000
Independent t-test						
A – B	***	***	n.s.	n.s.	***	***
A – C	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
B – C	***	***	n.s.	n.s.	***	*

[a] $n = 10$.
[b] $n = 9$.
[c] $n = 5$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 7 Correlations between the number of words that were used in three, two or one delivery in a training session and the gain score from pretest to immediate posttest for each fluency measure

		Mean length of fluent runs	Phonation/ time ratio	Mean length of pauses	Articulation rate
Session A	3 deliveries	0.335	0.413*	−0.461*	0.375
	2 deliveries	0.216	0.201	−0.107	0.024
	1 delivery	−0.365	−0.504*	0.521**	−0.274
Session B	3 deliveries	0.275	0.372	−0.444*	0.395
	2 deliveries	0.227	0.243	−0.277	0.312
	1 delivery	−0.177	−0.396	0.511*	−0.218
Session C	3 deliveries	0.382	0.545**	−0.532**	0.335
	2 deliveries	0.354	0.470*	−0.424*	0.161
	1 delivery	−0.328	−0.519**	0.590**	−0.307

Note. $n = 24$.
*$p < .05$.
**$p < .01$.

These correlations indicate that students who repeated more words across all three deliveries showed greater improvement, in that they were able to fill more time with speech and have shorter pauses on the immediate posttest than on the pretest. On the other hand, students who used more words in only one delivery showed a smaller gain in fluency from pretest to posttest in terms of phonation/time ratio and pause length. Correlations between the number of repeated (and nonrepeated) words and mean length of fluent runs and articulation rate were not significant.

It could be expected that the repeated words were those that were specifically related to the topic. For example, students speaking about sports are likely to use words like *sports*, *football*, and *score* in all three deliveries. To examine this, the number of topic-specific words per delivery was counted. We define topic-specific words as words that have a clear semantic relationship to the topic, in that they can be expected to be used with mostly that topic and less so with another topic in this study. It can be expected that students use the words *soccer* and *play* when talking about sports but not shopping. On the other hand, *favorite* can be used for both topics (e.g., *my favorite sportsman, my favorite store*) and is therefore considered non-topic-specific. Proper names of people and organizations were included in the analysis. Table 6 shows that some of the repeated words were specific to the topic, but, more importantly, most were not.

For example, student #164 from the Repetition condition used the following words in two or three deliveries in Session A. The words that can be considered as specific to the topic (Sports) are underlined.

Words used in all three deliveries (student #164)
favorite (Adj), even (Adv), very (Adv), also (Adv), well (Adv), Beckham (proper name), David (proper name), soccer (N), sport (N), sportsman (N), TV (N), be (V), know (V), like (V), make (V), play[4] (V), prefer (V), watching (verb) [18 words, of which 4 were topic-specific]

Words used in two out of three deliveries (student #164)
famous (Adj), good (Adj), healthy (Adj), really (Adv), why (Adv), sometime (Adv), America (proper name), British (proper name), Cup (proper name), England (proper name), Europe (proper name), World (proper name), day (N), friend (N), game (N), man (N), partner (N), thing (N), village (N), do (V), feel (V), go (V), have (V), practicing (verb), remember (V), say (V), see (V), watch (V) [28 words, of which 5 were topic-specific]

Only 9 out of these 46 repeated words were specifically related to the topic of Sports. These observations are in stark contrast to the data of student # 286 from the No Repetition condition. Again, words specific to any of the three topics are underlined (Sports, Learning English, and Travel).

Words used in all three deliveries (student #286)
good (Adj), example (N), be (V), have (V) [4 words, of which none was topic-specific]

Words used in two out of three deliveries (student #286)
important (Adj), very (Adv), in (Adv), always (Adv), well (Adv), Colombia (proper noun), States (proper noun), United (proper noun), country (N), family (N), kind (N), time (N), know (V), make (V), prefer (V), think (V), travel (V), want (V) [18 words, of which 1 was topic-specific]

Only 1 of the 22 repeated words was specifically related to one of the topics. In contrast, the student from the Repetition condition was able to repeat 49 words, only 9 of which can be considered specific to the topic; the other 40 words are more general. Similar patterns were found for the other students and in the other sessions. In sum, these results indicate that students who were asked to repeat their speeches did so indeed, repeating many words across the three deliveries, but most of those words were not topic-specific.

At first sight, it appears that the increase in fluency found in the two Repetition conditions could be attributed to lexical retrieval, but we argue

Table 8 95% confidence intervals of the three measures of proceduralization on Test 1

	Mean length of fluent runs		Mean length of pauses		Phonation/time ratio		Articulation rate	
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
Repetition[a]	3.61	5.46	1.00	1.32	48.30	61.21	177.99	211.69
No Repetition[b]	3.45	5.39	0.71	1.04	52.53	66.14	176.67	211.15
Repetition-II[c]	3.28	5.89	0.72	1.17	48.17	66.42	183.81	230.07

Note. Mean length of runs is given in syllables; mean length of pauses is given in seconds; phonation/time ratio is given in percentage; articulation rate is given in syllables per minute.
[a] $n = 10$.
[b] $n = 9$.
[c] $n = 5$.

that this is not the case. Although those words that were used in all three deliveries may have been retrieved more easily during the posttest, on average, students in the two Repetition conditions used only six of the repeated words on the immediate posttest, which is just three more than in the No Repetition condition. Moreover, many of these words were high-frequency words such as *be, have, good, make,* and *think*. It can be speculated that it is not the words themselves but the processing of sentence constructions and expressions they are used in that was proceduralized. This calls for further analysis, but it is outside the scope of this article.

Progress on Speeches With Identical Topics

In order to track the students' progress on more comparable speeches, the topic on Test 1 was the same as the topic of Session C (for the No Repetition condition, only the first speech in that session was the same as on Test 1). Ninety-five percent confidence intervals were computed for each of the four measures on Test 1 (see Table 8). The values of those confidence intervals were then compared to the values of the fluency measures of the speeches in Session C.

A number of observations can be made. First, in the No Repetition condition, the mean length of fluent runs and phonation/time ratio on the first delivery of Session C (3.52 and 53.12%, respectively) were close to the lower bound of the confidence interval of Test 1 (3.45 and 52.53%, respectively).[5] Thus, there was a temporary drop in performance. This may be related to the fact that the first delivery in Session C was 4 min long, compared to 2 min on Test 1. It may be

difficult for learners to maintain a high level of performance in longer stretches of speech (cf. Skehan & Foster, 2005). Interestingly, such a drop was not found for the two Repetition conditions, indicating perhaps that it was canceled out by the benefit of the two intervening training sessions. A second observation is that in the Repetition condition, the mean length of pauses on all deliveries of Session C was below the lower bound confidence level of Test 1 (i.e., 0.99, 0.89, and 0.82 are below 1.00). This indicates an improvement in performance, possibly due to the two intervening training sessions. A third observation is that in the Repetition condition, the phonation/time ratio of the last delivery of Session C (61.71%) was above the upper bound confidence interval of Test 1 (61.21%). Therefore, it seems that the students in this condition were able to improve their performance during the session while repeating their speech and ultimately exceed their level of performance on Test 1. Finally, in the No Repetition and Repetition-II conditions, the articulation rates of the last delivery of Session C (216.85 and 238.56, respectively) were above the upper bound confidence interval of Test 1 (211.15 and 230.07, respectively). This indicates that the students in these two conditions were able to improve their performance during the session. In the case of the Repetition-II condition, this may be related to their repeating the speech. In the case of the No Repetition condition, it may be due to the difference in topic. The topic of the last speech was e-mail, which more students had a more favorable opinion of than pets. In sum, it seems that the students in the two Repetition conditions were able to benefit from speech repetition to improve their performance (e.g., reaching a substantive increase in the phonation/time ratio by the last delivery of the 4/3/2 task). In general, pets seemed to be an unpopular topic with the students in this study, as evidenced by informal feedback from the students and the content of their speeches, in which they often expressed negative opinions about pets. This may be a partial cause of the lower levels of performance on the fluency measures compared to speeches on other topics. Nevertheless, the drop in performance seemed smaller for the two Repetition conditions than for the No Repetition condition, which suggests an effect of the preceding 4/3/2 sessions.

Discussion

Hypothesis 1 was supported, in that the increases in fluency found in the pretests and posttests show evidence of proceduralization, but only in the two Repetition conditions. The Repetition-II condition showed the pattern as described by Towell et al. (1996): increased mean length of fluent runs with a stable mean length of pauses and phonation/time ratio. The Repetition condition showed a

different pattern: decreasing pause length, increasing phonation/time ratio, and a stable mean length of fluent runs. These two patterns may reflect alternative manifestations of proceduralization. Speakers may produce longer stretches of fluent speech without having to pause more for planning, as measured by the mean length of pauses and phonation/time ratio. Alternatively, speakers may produce stretches of speech of the same length but with less pausing for planning. Therefore, it can be concluded that the students in both conditions had proceduralized some of their L2 knowledge.

Hypothesis 2 was supported, in that fluency improvements were maintained over 4 weeks and transferred to new topics, but only in the two Repetition conditions. In the Repetition condition, pause length decreased and phonation/time ratio increased from pretest to immediate posttest, and this was maintained in the delayed posttest. In the Repetition-II condition, mean length of fluent runs increased from the pretest (Time 2) to the immediate posttest (Time 3). This condition did not have a delayed posttest. In contrast, performance under the No Repetition condition did not change significantly during the semester. Interestingly, articulation rate (in syllables per minutes) increased mostly from Test 2 to Test 3. However, because this group had not received the fluency training between these two tests, this increase in articulation rate cannot be ascribed to the training. Instead, it may have been a result of the continuing Speaking classes. In sum, these results from the pretests and posttests show that cognitive fluency increased, but only as a result of speech repetition in the training.

It is important to note that the improvements described here took place mostly between the pretest and immediate posttest and can thus be ascribed to the 4/3/2 fluency training. In addition, the fluency measures in the Repetition condition increased between Test 1 and Test 2 (their pretest and immediate posttest) and were stable between Test 2 and Test 3, whereas in the Repetition-II condition, the fluency measures were stable between Test 1 and Test 2 (their two pretests) and increased between Tests 2 and 3 (their second pretest and immediate posttest); that is, there were improvements only during the period in which students received the fluency training. In addition, it should be stressed that these effects were found on posttests that were administered 1 and 4 weeks after the last training session and involved new topics. The effect of speech repetition, therefore, went beyond the training sessions themselves.

In examining the source of the gains, it was found that performance in all three conditions improved on all four measures during the sessions; however, performance in the Repetition conditions did not improve over and above that of the No Repetition condition. It may be speculated that the topics of the speeches were of influence, as the improvements appear to be steadier in

the two Repetition conditions (Repetition and Repetition-II) than in the No Repetition condition with varied topics. Alternatively, gains may reflect not proceduralization, but momentary effects of time pressure.

One explanation for the differences in longer term fluency gains between the Repetition and No Repetition groups was revealed by an analysis of lexical overlap, which examined the number and types of words repeated across deliveries in a training session. There were two notable results. First, students who repeated more words across deliveries showed greater improvement from pretest to posttest in terms of the phonation/time ratio and length of pauses. Second, the analysis of lexical overlap showed that not only the number of repeated topic-related words was higher in the two Repetition conditions but also, and more importantly, the number of repeated non-topic-related words. Moreover, only few repeated words were also used in the immediate posttest (five or eight from each session in the two Repetition conditions and three or four in the No Repetition condition). It seems likely, therefore, that proceduralization was not a specific lexical effect; rather, the effect may have been in the repeated use of sentence structures with those repeated words, thus leading to proceduralization of phrase building. Further analysis is needed to find out if sentence structures are repeated, what types of structures, and if they contribute to longer fluent runs and shorter pauses.

In sum, performance fluency and cognitive fluency seemed to have improved from pretest to posttest in the two Repetition conditions (i.e., after training in which students repeated their speeches). In contrast, performance in the No Repetition condition improved during the training sessions, but this improvement was not retained in the posttests. The effect found in the Repetition conditions may be ascribed to the proceduralization of linguistic knowledge due to repeated use. The analysis of the overlap between speeches in terms of vocabulary showed that the improvements were not limited to vocabulary related to the topics in the training but to more general vocabulary. A similar effect may be expected to have taken place for morphological and syntactic structures, but that requires further investigation.

Overall, the present study shows that repeated practice increases fluency. In itself this is not a new finding, but it is nontrivial for several reasons. First, transfer and long-term retention were found. The transfer of improvements to new topics shows that the results cannot be explained by lexical priming, as much of the same vocabulary could not be used. Indeed, it was found that only few words used in the 4/3/2 task were also used in the posttests. In addition, there was a long-term effect. During the 4/3/2 task itself, across deliveries, fluency may increase due to lexical and structural priming: When vocabulary and

grammatical structures are more readily available, fewer searches are needed, reducing the number and length of pauses. Furthermore, fluency may increase due to planning and attentional resources: Because the students know what to say and how to say it, they have more resources for retrieving vocabulary and grammatical structures, again reducing the need for frequent and long pauses. However, in the present study, the long-term effect cannot be explained by priming or planning, as there was a delay of several weeks and there were no differences in planning between the conditions. The effect, therefore, must be attributed to changes in the students' underlying knowledge and processing.

Such long-term retention and transfer to new topics have not been shown in previous research. Many studies of fluency, such as Nation's (1989) and Arevart and Nation's (1991) studies of the 4/3/2 task, did not include immediate and delayed posttests. Bygate (2001) did not find transfer to a new topic. However, his study did not include repeated practice, so it is unlikely that there were changes in underlying processing mechanisms that could have had a broader effect than planning a particular speech. Often, transfer does not take place from one task to another because proceduralized knowledge is highly specific and a common component between tasks is missing. In the present study, however, the tasks were similar in that they all required oral production of monologues, albeit about different topics. The common component between the training and the posttests therefore may have mostly been of a morphological or syntactic, rather than lexical, nature.

A second reason for the importance of our findings is that the effect of repetition seems to scale up from item and sentence level practice to longer stretches of speech of up to 4 min. Effects of repetition at the item or sentence level have been shown in many studies, both for language learning and other types of learning (e.g., Anderson et al., 2004; Anderson & Lebiere, 1998; De Jong, 2005; DeKeyser, 1997; Ferman et al., 2009). Gatbonton and Segalowitz (2005) argued that inherently repetitive but communicative activities can promote automaticity. This study shows that the 4/3/2 task is one such activity. Although in this study students spoke to a computer, in the original 4/3/2 task students spoke to three different classmates, which is more naturalistic and communicative.

The repetition of speeches in the 4/3/2 task is likely to have led to changes in the underlying knowledge and processing mechanisms and cannot be explained as faster retrieval due to the repeated use of specific lexical items. More likely, changes affected the encoding stages of language production, like phrase and clause structure building (cf. Kormos, 2006; Levelt, 1999). Students may have been able to form new production rules and strengthen them by repeated use

(Anderson et al., 2004; Anderson & Lebiere, 1998). This could show up as repeated use of certain phrases and phrase structures and perhaps formulaic sequences (Towell et al., 1996; Wray, 2002). Proceduralization is considered a slow process that requires many encounters with the same items. However, in the present study, words and phrases were repeated relatively few times. Therefore, it seems the improvement in performance might be a reflection of the initial stages of proceduralization, in which new *production rules* are formed that lead to relatively greater gains in performance. Thus, an important question remains unanswered: Although there was evidence that proceduralization of language knowledge took place, it is not clear exactly what knowledge was proceduralized. It was argued that more than just topic-related vocabulary was involved. A deeper, qualitative analysis of the types of grammatical structures used and perhaps the emergence of formulaic sequences will give indications of what knowledge was proceduralized, and a detailed analysis of syntactic complexity and accuracy can assess if there was a trade-off among accuracy, complexity, and fluency or if the 4/3/2 procedure in fact led to higher accuracy and complexity. However, such analyses would deserve an in-depth discussion that is outside of the scope of the present article; a separate study is currently in progress. Finally, future studies will need to include focused tests of vocabulary and grammar before and after training to identify where development takes place.

Conclusion

This study not only investigated fluency development but also examined underlying changes in the processing of language knowledge. In addition, it combined data from training tasks and pretests and posttests, in order to study long-term effects and to clarify the causes of the increase in fluency. It was shown that fluency increased during the 4/3/2 task, in which students spoke three times, for 4, 3, and 2 min, respectively. However, this increase in fluency only transferred to a speech about new topic when the students had repeated their speeches in the 4/3/2 training. More importantly, their improvement was most likely due to proceduralization because after the training they were able to produce fluent runs of similar lengths but filling more time with speech and pausing less long. This proceduralization may have been due to the repeated use of particular words and sentence structures because it was found that those students who repeated more words across the three deliveries showed higher gains in fluency from pretest to posttests, even though few of these repeated words were semantically related to the topic. In addition, very few of these

repeated words were used again in the immediate posttest. Proceduralization, therefore, clearly concerned more than the retrieval of topic-related words. In conclusion, speech repetition in the 4/3/2 task may cause changes in underlying cognitive mechanisms, resulting in a long-term and transferrable effect on performance fluency. Having established the overall effect of repetition on fluency development, a deeper, qualitative analysis of the students' production is needed.

Revised version accepted 6 November 2009

Notes

1 Proceduralization of sentence building may in part involve the use of formulaic sequences, such as *the point is that* and *to give an example*. At this point it is not clear how many such formulaic sequences were used in these data, especially because many of the sequences may be idiosyncratic; few nativelike sequences seem to have been used.

2 Although the students in this institute are in an immersion setting—in that the target language is the dominant language and all classes are taught in the target language—a large part of the students' language learning takes place in the classroom. In addition, due to the design of the study, the effect of the training can be isolated, and results are expected to be generalizable to nonimmersion classroom settings. Swain (1991) found that students in a French immersion setting in Canada had limited opportunities to engage in oral production and much of their "public" talk was not longer than a clause. In contrast, the 4/3/2 task provides the students with practice in monologues of up to 4 min.

3 The [/], [//], and [///] symbols indicate the type of retracing, and the < and > symbols indicate stretches of speech that were retraced.

4 Some words can be seen as related to other topics in the study. These were examined in the other speeches as well. To give an example, the verb *play* was also used in Session B by two students and in Session C by six students, all of whom used it in only one delivery. In addition, in Session C it was only used with the Pets topic. In comparison, the same verb was used in Session A by 19 students, 10 of whom used it in all three deliveries and three used it in two deliveries. Therefore, we can conclude that the verb *play* is more strongly semantically related to Sports than to any other topic in this study. Importantly, the verb *play* was used by only four students in Test 1 (Pets), two in Test 2 (Important Person), and none in Test 3 (Biggest Problem) and thus confirms that topic-specific words did not have the strongest effect on the fluency measures in the tests.

5 In fact, in the No Repetition condition the mean of the mean length of fluent runs on Test 1 (4.42) is above the upper bound of the 95% confidence interval on the first speech of Session C (4.18), indicating that performance on the 4/3/2 task is below performance on Test 1.

References

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036–1060.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

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

Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition, 35*(1), 1–39.

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341–345.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*(2), B13–B25.

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

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (Vol. 11, pp. 37–74). Amsterdam: Benjamins.

Chambers, F. (1997). What do we mean by fluency? *System, 25*(4), 535–544.

Corrigan, A., Dobson, B., Kellman, E., Spaan, M., Strowe, L., & Tyma, S. (1979). *Michigan Test of English Language Proficiency*. Ann Arbor: University of Michigan.

De Jong, N. (2005). Can second language grammar be learned through listening? An experimental study. *Studies in Second Language Acquisition, 27*(2), 205–234.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19*, 195–221.

DeKeyser, R. M. (2007). Study abroad as foreign language practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 208–226). Cambridge: Cambridge University Press.

Ferman, S., Olshtain, E., Schechtman, E., & Karni, A. (2009). The acquisition of a linguistic skill by adults: Procedural and declarative memory interact in the learning of an artificial morphological rule. *Journal of Neurolinguistics, 22*, 384–412.

Foster, P., & Skehan, P. (1996). The influence of planning and task type of second language performance. *Studies in Second Language Acquisition, 18*, 299–323.

Freed, B. F. (1995). What makes us think that students who study abroad become fluent? In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123–148). Amsterdam: Benjamins.

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and

intensive domestic immersion programs. *Studies in Second Language Acquisition, 26*, 275–301.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review, 61*(3), 325–353.

Goldman-Eisler, F. (1961). The distribution of pause durations in speech. *Language and Speech, 4*(4), 232–237.

Hilton, H. (2008). Connaissances, procédures et production orale en l2. *AILE, 27*, 63–91.

Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 5–24). Ann Arbor: University of Michigan Press.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Erlbaum.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*, 145–164.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*(3), 387–417.

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.), *The neurocognition of language* (pp. 83–122): Oxford: Oxford University Press.

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling in classroom language learning. *Language Teaching Research, 4*(3), 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: Second language learning, teaching and testing* (pp. 141–162). Harlow, UK: Pearson Longman.

Maurice, K. (1983). The fluency workshop. *TESOL Newsletter, 17*(4), 29.

McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning, 56*(4), 693–720.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition, 20*, 83–108.

Nation, P. (1989). Improving speaking fluency. *System, 17*(3), 377–384.

Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences, 3*(4), 136–141.

Schmidt, R. W. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14*, 357–385.

Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 200–219). Ann Arbor: University of Michigan Press.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition. *Studies in Second Language Acquisition, 26*, 173–199.

Shafer, D. (2006). *Revolution: Software at the speed of thought*. Monterey, CA: Shafer Media.

Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 193–216). Amsterdam: Benjamins.

Snellings, P., Van Gelderen, A., & De Glopper, K. (2002). Lexical retrieval: An aspect of fluent second language production that can be enhanced. *Language Learning, 52*(4), 723–754.

Snellings, P., Van Gelderen, A., & De Glopper, K. (2004). The effect of enhanced lexical retrieval on second language writing: A classroom experiment. *Applied Psycholinguistics, 25*, 175–200.

Squire, L. R. (1987). *Memory and brain*. New York: Oxford University Press.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 2*, 195–231.

Swain, M. (1991). Manipulating and complementing content teaching to maximize second language learning. In R. Philipson, E. Kellerman, L. Selinker, M. Sharwood Smith, & M. Swain (Eds.), *Foreign/second language pedagogy research* (pp. 234–250). Clevendon, UK: Multilingual Matters.

Towell, R. (2002). Relative degrees of fluency: A comparative case study of advanced learners of French. *IRAL, 40*, 117–150.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17*(1), 84–119.

Ullman, M. T. (2001a). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research, 30*(1), 37–67.

Ullman, M. T. (2001b). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews: Neuroscience, 2*, 717–726.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*, 231–270.

Wood, D. (2001). In search of fluency: What is it and how can we teach it? *Canadian Modern Language Review, 57*(4), 573–589.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Youjin, K., & McDonough, K. (2008). Learners production of passives during syntactic priming activities. *Applied Linguistics, 29*(1), 149.

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

Appendix

Topics
Tests

Test 1: a. How do you feel about pets? Do many people have pets in your country? How are they treated, in general? [Note: This topic is the same as the first topic in Session C]

Test 2: b. Talk about a person who was very important to you in the past. Who was this person? Why was this person important to you?

Test 3: c. What is the biggest problem your country is facing today? How would you change it?

Training (topics in bold were given in both conditions):
Session A

1. **Do you like sports? Why? If you do, what is your favorite sport? Why? Do you prefer watching the sport or doing it yourself? Who is your favorite sportsman or sports woman? Give an example of a game in which he or she played well.**
2. Do you think it is important to learn English? Why? Give an example of a situation in which English is important. Are other languages important for you? Which languages do you speak? What other languages would you like to learn? Why?
3. When you travel, what kind of transportation do you use? How do you prefer to travel if you have a choice? Does distance make a difference? Give an example of transportation you use for short and long distances.

Session B

4. **Do you like shopping? Why? What do you think of shops in the U.S.? Do you like them? Why? Can you buy everything you want? Give an example of something from your country that you can't buy in the U.S.**
5. What do you think about cell phones? Do you think they are useful? Give an example of why they are useful or not useful. How are cell phones used in your country?
6. What do you think about television? What do you like about it? What don't you like? Give an example of something you like and something you don't like about television.

Session C

7. **How do you feel about pets? Do many people have pets in your country? How are they treated, in general? [Note: This topic is the same as the topic in Test 1]**
8. What kind of clothing do you usually wear? Why do you like it? Is clothing in the U.S. different from clothing in your own country? How? Give an example of clothing that is different in your country.
9. What do you think of e-mail? Is it a good way to keep in touch with your family and friends? Do you prefer e-mail, phone, or letters? Why? Give an example of when you use e-mail.