The Relationship Between Task Difficulty and Second Language Fluency in French: A Mixed

Methods Approach

YVONNE PRÉFONTAINE
University of Illinois at Urbana–Champaign
Department of French
Urbana, IL 61801
USA
Email: yvonneprefontaine@yahoo.com

JUDIT KORMOS
Lancaster University
Department of Linguistics and English Language
Bailrigg, Lancaster LA1 4YL
United Kingdom
Email: j.kormos@lancaster.ac.uk

<ABSTRACT>

While there exists a considerable body of literature on task-based difficulty and second language

(L2) fluency in English as a second language (ESL), there has been little investigation with

French learners. This mixed methods study examines learner appraisals of task difficulty and

their relationship to automated utterance fluency measures in French under three different task

conditions. Participants were 40 adult learners of French at varying levels of proficiency

studying in a university immersion context in Québec. Appraisal of task difficulty was assessed

quantitatively by participants' self reports in response to a five-item questionnaire and

qualitatively by retrospective interviews. Utterance fluency was operationalized by four temporal

variables and measured by Praat, a speech analysis software program. Across tasks, the

quantitative results indicate that appraisals of lexical retrieval difficulty and fluency difficulty

were most strongly related to perceived overall task difficulty. The qualitative analysis shows

how L2 speakers evaluated the difficulty of each task as well as the features that either

contributed to or limited their L2 fluency. Students' fluency in performing the three tasks was

found to differ for articulation rate and average pause time, but not for pause frequency or phonation–time ratio.

Empirical models theorizing the nature of speech production (de Bot, 1992; Kormos, 2011; Levelt, 1989) describe how second language (L2) speakers involved in communicative tasks encode or formulate ideas into speech. Although there exist numerous points of *fluency vulnerability* (Segalowitz, 2010) in the phases from intention to articulation that have the potential to impair L2 fluency, little attention has been paid to the speech task itself in the French language learning literature. Recent research has studied French learners from various perspectives, including temporal variables to measure utterance fluency (French & Guay, 2013; Préfontaine, 2013a, 2013b; Towell, 2002; Towell, Hawkins, & Bazergui, 1996) and speech features that influence raters' evaluation of perceived fluency (Freed, 2000) in different learning conditions. But task-related variation and its interrelationship with speech production using automated measures has been a neglected component of empirical L2 fluency investigations (for a recent study with Dutch learners, see De Jong et al., 2012). Moreover, previous research findings on L2 task difficulty are primarily concerned with English as a second language (ESL) or foreign language (EFL) learners (Foster & Skehan, 1996; Ishikawa, 2011; Robinson, 2001, 2003, 2005; Robinson & Gilabert, 2007; Skehan & Foster, 1997, 1999, 2001; Tavakoli, 2009). This leaves unexplored the question of how task characteristics and learners' appraisals of the difficulty of tasks interact with speech performance in French.

This study intends to bridge this gap by using a mixed methods design. The novelty of our research is that it investigates appraisals of task difficulty, that is, learners' perceptions of the ease or difficulty involved in performing a given task, and their relationship to fluency from both quantitative and qualitative perspectives; it does so by applying automated utterance fluency measures, post-task questionnaires, and retrospective interviews. This approach is motivated by the nearly singular adherence to quantitative methods in previous studies evaluating the

relationship between different task characteristics and temporal correlates of speech. Further, such research should expand our understanding of how task characteristics are linked to psycholinguistic processing and speech production.

<A>THE CONSTRUCTS OF TASK COMPLEXITY AND TASK DIFFICULTY

Task complexity and task difficulty are key constructs in studies of task-based learning and assessment (see, for example, Bachman, 2002; Robinson, 2001; Skehan, 1998). The importance of these constructs lies in the fact that, in task-based language teaching (TBLT), the particular characteristics of tasks that learners perform are assumed to drive linguistic development (for a review, see Samuda & Bygate, 2008). Understanding how the requirements of tasks foster learning informs the sequencing of tasks in task-based syllabi (Robinson, 2011) and the design of communicative tests of language ability (Norris et al., 1998). To date, most of the studies in this field have focused on the concept of task complexity, which is perceived to be an inherent characteristic of the task independent of the student(s) performing the task and the context in which it is performed. While there is no doubt that understanding this aspect of task complexity is instrumental to the field of second language acquisition (SLA) and TBLT, it should also be remembered that students often have their own linguistic and communicative agendas in relation to task performance, which might not align with those of researchers and task designers (see, for example, Coughlan & Duff, 1994; Mori, 2002). Therefore, it is necessary to investigate learners' perceptions of tasks and their views concerning various task features (for a similar argument, see Révész, 2014). Learners' appraisals of task difficulty can potentially provide insights into some of these issues. An understanding of perceptions of difficulty can also help task designers, teachers, and language testers to devise tasks for learners that provide an

optimal level of challenge and thereby enhance the motivation to perform tasks.

In the field of task-based research there exist two different conceptualizations of task difficulty. In his Cognition Hypothesis, Robinson (2001) defines task difficulty as "the result of the attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner" (p. 28). In his model, task difficulty pertains to learners' perceptions of task demands and factors that contribute to differences between learners in their production, such as ability, aptitude, confidence, or motivation. Specifically, perceptions of task difficulty are influenced by the learner's ability and affective factors. He argues that task complexity, which encompasses the cognitive and linguistic demands of tasks, needs to be separated from both learners' perceptions of difficulty and task conditions, as they determine how a particular task is administered (e.g., interactively or individually). In contrast, Skehan (1998) views task difficulty as an overarching concept that includes code complexity, in other words the linguistic demands of tasks, communicative stress, and cognitive complexity.

Particularly relevant to the discussion of how task difficulty and task complexity can be conceptualized is Tavakoli's (2009) study which investigated the perceptions of task difficulty of ten L2 learners and ten teachers using four narrative tasks in an EFL/ESOL context. Using retrospective interviews, the study sought to determine why some learners and teachers found certain tasks more difficult than others and what were the contributing factors. The research revealed that learners' and teachers' perceptions were largely similar with regard to variables such as cognitive and linguistic demand, the amount of information needed to complete a task, and task structure. In other words, teachers' and learners' evaluations of task complexity in the sense defined by Robinson (2001) were aligned. Nevertheless, learners' perceptions differed from those of teachers with reference to the factors to be considered to determine task difficulty

and the perceptions of interesting tasks. Tavakoli concluded that it was problematic to relate learners' and teachers' criteria for task difficulty to Robinson's framework, because he considers cognitive factors to be either *resource directing* or *resource dispersing*, which is troublesome, as tasks can direct learners' attention to particular linguistic features while at the same time making demands on the availability of their attentional resources. Tavakoli argued that the findings could be more easily interpreted in relation to Skehan's model because it takes into consideration cognitive factors, such as cognitive demand, linguistic demand, the clarity of pictures/stories, the amount of information, task structure, and affective factors, which also affect task difficulty.

Task difficulty is also an important construct in task-based language assessment as it helps test designers provide candidates with appropriate tasks (e.g., Brindley & Slatyer, 2002; Norris et al., 1998). Nevertheless, establishing the link between the linguistic and cognitive demands of communicative test tasks at the design stage and students' actual performance on the task has proved to be rather difficult, and this can also be seen in the conflicting results obtained in task-based studies (for a meta-analysis, see Jackson & Suethanapornkul, 2013). Bachman (2002) points out that most of the task characteristics assumed to constitute the cognitive complexity of tasks need to be considered in interaction with learners' ability and other individual characteristics, as well as the communicative context. He further argues that "difficulty is not a separate factor at all, but resides in interactions" (p. 466) with the aforementioned factors.

Bachman (2002) makes several recommendations for future research that are worth considering in the field of task-based research. First, he suggests that tasks should be "conceptualized as sets of characteristics rather than as holistic entities" (p. 469). Given the inconclusive findings of a vast number of research studies that compared complex vs. simple

tasks (see, for example, De Jong et al., 2012; Gilabert, 2007; Robinson, 1995, 2001, 2003, 2005, 2007; Robinson & Gilabert, 2007; Skehan & Foster, 1997, 2001; Tavakoli & Skehan, 2005), we might gain more detailed insights into task-based performance and learners' task-related cognitive processes if we give careful consideration to the particular demands of tasks. Skehan's (2009) recent discussion of task complexity might be relevant in this respect as it makes an attempt to relate various task characteristics to Levelt's (1989) model of speech production. Skehan proposes that one needs to consider both the influence of cognitive demands on the conceptualization of the speaker's message and the demands that the task content, structure, and set up make on syntactic and lexical encoding (see Kormos & Trebits, 2012, for a similar line of argumentation). Bachman (2002) also recommends that researchers give careful consideration to establishing which task features and demands are inherent to the task and which ones interact with the attributes of learners. These suggestions, however, can only be followed if, in addition to examining how the manipulation of various task features impacts on performance, information is collected on learners' cognitive and affective appraisals of tasks and cognitive processing and strategic behavior during task performance.

<A>CONSTRUCTS AND OPERATIONALIZATION OF L2 FLUENCY

Empirical findings about how specific task characteristics impact on fluency in task performance differ significantly, which is in part due to varying operationalization of fluency. While most task-based studies considered fluency in its narrow sense, that is, as it refers to the "rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention under the temporal constraints of on-line processing" (Lennon, 2000, p. 26), rather than as a measure of global proficiency in its broad sense (see Fillmore, 2000), research has used different

variables to assess fluency. This variation stems from different conceptualizations of utterance fluency, in other words, from the variables that can be used to describe 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 measured as: *breakdown fluency*, which is related to pausing behavior, *repair fluency*, which is operationalized as the frequency of repetitions and self-corrections, and *speed fluency*, which expresses the speed with which speech is delivered (for a recent discussion, see Bosker et al., 2013; De Jong et al., 2012). To illustrate, Robinson (2001) operationalized fluency in task performance as the number of words per communication unit (C unit)—which consists of an independent clause with its modifiers—and as pauses per utterance and words per utterance (Robinson, 2005)—which is a combination of speed and breakdown fluency. Skehan and Foster (1997) took a different approach and complemented these measures with variables assessing repair fluency and hence analyzed the number of reformulations, replacements, false starts, and repetitions. As Jackson and Suethanapornkul's (2013) synthesis reveals, the 17 studies reviewed applied 17 different measures, many of which have not been found reliable in the assessment of fluency (De Jong et al., 2012; Kormos & Dénes, 2004; Towell et al., 1996)

The operationalization of fluency in our study is based on the work of De Jong et al. (2012), who give a precise description of the utterance fluency variables that can be used to assess different facets of fluency. De Jong et al. (2012) explain that pause frequency, pause length, and phonation–time ratio—the ratio of the total length of time spent speaking and total utterance time—are the most accurate measures of breakdown fluency. Based on the work of Goldman–Eisler (1968), they suggest the use of articulation rate, that is, the number of syllables divided by speaking time, to measure speed fluency. In their study they also investigated repair

fluency by analysing the frequency of self repairs in L2 learners' speech. But because this aspect of oral performance might also reflect learners' attention to the accuracy of their output (Gilabert, 2007; Kormos, 2000) it was not included in our study. Using these measures, De Jong et al.'s (2012) study found that articulation rate was not sensitive to variation across simple and complex tasks, but in a complex task, L2 learners of Dutch had a lower phonation–time ratio and greater frequency of filled pauses.

Despite the large number of studies investigating L2 speech production, few have addressed how task characteristics relate to fluent speech performance using a set of utterance fluency measures, questionnaires assessing learners' perceptions of task difficulty, or retrospective interviews. Of these investigations, none has been conducted with French language learners. Moreover, previous research has analysed the relationship of task characteristics with only a small set of utterance fluency measures (see De Jong et al., 2012, as an exception), and qualitative feedback on perceptions of task difficulty in French learners is negligible in the SLA literature. Thus, based on the considerations just described, the main goal of this study was to examine at some depth learners' perceptions of task difficulty and the relationship of these appraisals to L2 fluency in French, using both qualitative and quantitative data collection methods. Consequently, the present research was guided by four main research questions:

RQ1.   What quantitative and qualitative differences exist in the perceived difficulty of narrative tasks in French as an L2?

RQ2.   How are perceptions of task-related difficulty in various areas of linguistic processing related to appraisals of overall task difficulty?

RQ3.   How does L2 utterance fluency in French vary across three types of speech tasks?

RQ4.   How is the perceived difficulty of tasks related to utterance fluency in the speech of L2

learners of French?

<A>METHOD

<B>*Participants*

The L2 speakers were 40 undergraduate and graduate students with varying levels of

French proficiency who were enrolled in a full time five-week French immersion programme at a

university in Québec, Canada. The selected participants were all volunteers from an L1 English-

speaking background and were between 18 and 69 years old ($M = 26$ years, $SD = 10.57$). The

sample population was intentionally heterogeneous to allow for variation in L2 fluency in

French. There were two reasons for this decision: first, to obtain a wide spread of fluency scores

needed for correlation in order to avoid having a ceiling effect, and second, to examine the

characteristics of tasks that contribute to perceptions of difficulty and to be able to generalize

about perceptions of task difficulty regardless of proficiency level. The participants were drawn

from five levels of language proficiency on the basis of an institutional placement test in the four

skills (pre-intermediate [$n = 8$], intermediate [n = 9], upper intermediate [$n = 7$], advanced [$n =$

9], and proficient [$n = 7$]). The participants included 26 Canadians, 13 Americans, and one

British student, of whom 21 were females and 19 were males. With the exception of 10 Canadian

participants who had an average of nine years in a French immersion program in a province other

than Québec, the group had an average of six years of French language learning experience in a

classroom setting.

<B>*Instruments*

<C>*Speech Tasks*. Cognizant that L2 utterances and perceived fluency vary according to the various demands posed on different aspects of speech production, namely the conceptualization of the message and giving it linguistic form, the study sought to employ three different narrative task types and degrees of task difficulty. We selected three types of narratives that varied systematically with regard to the need to invent a storyline, and thereby posed different demands on learners' creative, content-related, and linguistic resources.

The task that required the highest level of creativity, and consequently was assumed to be the most taxing for the conceptualization stage of speech production, was unrelated picture narration (Task 1) that required L2 speakers to construct a storyline from a set of six unrelated pictures. In this task, students had to invent their own story involving these six pictures and narrate it. Although this task required the generation of content on the part of the learner, it also allowed for matching the content of the story with the learners' existing linguistic resources and avoiding syntactic constructions or lexical items that the learners could not express in the L2. This task was selected as it included features that can potentially both limit and enhance L2 fluency and evoke varied appraisals of task difficulty of the different processes involved in performing the task.

In contrast, in the story retelling task (Task 2), L2 speakers were asked to retell a horse-riding accident based on a short text in English. This task required retelling a story from memory that learners had first read in their L1. It was considered to require the mobilization of several different verb tenses and context-dependent vocabulary, thus making increased demands on the formulation stage of production. The researchers made sure the participants understood that the task was not an exercise in translation. Rather, the goal was simply to tell the story in French as if they had read it in the newspaper and wanted to tell someone about it. Although this type of

task is uncommon in second language research studies, the event of reading a story in one's L1 and telling it in an L2 is not uncommon in daily life. In comparison with the unrelated picture narration task, this task did not require the creative conceptualization of a storyline and left little room for the learners to adjust the message to their linguistic resources and avoid constructions of which they might have limited knowledge. Consequently, we hypothesized that this task would prove to be difficult for the learners in general, as well as with regard to the linguistic encoding stage of speech production.

We selected a third task in which L2 speakers narrated a story based on an 11-frame cartoon strip presented in sequential order. In this related comic strip picture task (Task 3), the conceptualization demands were lower than in Task 1, in that the storyline was given and the sequence of events to be narrated was predictable. For this task, however, participants had somewhat more opportunities to tailor the message to their existing linguistic resources than in the story retelling task (Task 2), because they could make decisions on the level of detail they gave in their descriptions. For this reason, we assumed that this task would be perceived as the least difficult of the three, in terms of both conceptualization and formulation demands.

<C>Perceived Task Difficulty Questionnaire. A task difficulty questionnaire was administered in English to assess the participants' overall perceptions of speaking difficulty across tasks. It was adapted from Robinson's (2012) five-item scale and included items related to perceptions of task difficulty as used in stages of speech production, namely, speech planning, lexical retrieval, and grammatical encoding, as well as perceived difficulties in speaking fluency (see Appendix A). The reliability of the questionnaire was high (Task 1 Cronbach's alpha = .81; Task 2 Cronbach's alpha = .90, Task 3 Cronbach's alpha = .93).

<C>Retrospective Interviews. These were conducted according to the guidelines set out by Ericsson and Simon (1980, 1993). The goal of retrospective interviews was to capture and explore participants' thoughts by attempting to tap into information about task difficulty relative to speech processing retained in memory. Immediately after the three speech performances were completed, the participants provided justifications for their perceptions of task difficulty in English. This open ended approach allowed the participants to vocalize their own judgements about what constitutes task difficulty for them personally, and also to comment on the specific features of tasks, which they believed affected their speech performance.

<C>Online Placement Test. The online multiple choice placement test scores identified students by language ability in terms of three distinct components: Grammar (54 items), Vocabulary (30 items), and Listening Comprehension (16 items). The fourth component consisted of an oral interview, lasting from 10 to 20 minutes, with a faculty member from the language school, followed by a writing task.

<B>Procedure

The researchers and participants worked together on a one-on-one basis in the French immersion department at the university. The placement test scores for the participants were provided for the purposes of this research by the institution. At the data collection meeting, which lasted approximately one hour, participants responded to three different speech tasks, with a three-minute planning allotment per task. Individually, the participants looked at the speech elicitation instruments and self-initiated their narratives upon signalling to the researchers when they were ready to begin speaking. Even though the participants had different levels of French competency, none used the entire planning time allotted. During task completion, the students were allowed to view the instruments. The participants were informed that there were no correct

or incorrect responses and were instructed to speak for as long as they felt natural to complete the task, which prompted renditions of between one and four minutes. The tasks were administered in a counter-balanced design to control for task order effects. Immediately after each task performance, participants completed the Task Difficulty Questionnaire, followed by a retrospective interview consisting of one closed question about which task they perceived to be the most or least difficult in terms of fluent speech production and one open ended question regarding the task characteristics they believed contributed to this perception. All speech samples and task difficulty comments were recorded in Apple GarageBand version 6.0.1. The L2 speech performances were converted to audio files so that they could be automatically evaluated using Praat, a speech analysis software program designed at the University of Amsterdam (Boersma & Weenink, 2010; see http://www.fon.hum.uva.nl/praat/).

<B>*Analyses*

Perceived task difficulty was measured both quantitatively and qualitatively in the following two ways. First, the participants rated each of the three task performances by means of the Task Difficulty Questionnaire. For the analysis, the task difficulty responses were computed on a Likert scale ranging from one to six, where one represented the lowest level of difficulty and six the highest. Second, the participants' retrospective interview comments were analysed to identify the main task characteristics involved in perceptions of task difficulty and features enabling or disenabling fluent performance. The interviews were recorded and transcribed. Following Rubin and Rubin's (2005) guidelines, we then extrapolated the concepts that represented the two retrospective interview questions and identified themes. Next, we synthesized these themes and selected topics pertinent to the research question. Finally, the data were coded and summarized according to how the learners' contextualized the varying themes.

The first author of the article carried out the coding, which was then verified and double coded by the second author.

Utterance fluency was analysed by calculating articulation rate, pause frequency, phonation–time ratio, and average length of pauses. Following De Jong et al. (2012), we selected articulation rate to assess the construct of speed fluency and pause frequency, and average length of pauses and phonation–time ratio as measures of breakdown fluency. These fluency variables were then examined using Praat and a software script written by De Jong and Wempe (2009), which was modified by the authors to extract automatically the temporal measures listed. These utterance fluency measures were also chosen for inclusion in the analysis because previous empirical research has reported that they are valid fluency predictors (Bosker et al., 2013; Derwing et al., 2009; Derwing et al., 2004; Ejzenberg, 2000; Freed, 2000; Freed, Segalowitz, & Dewey, 2004; Iwashita et al., 2008; Kormos & Dénes, 2004; Lennon, 1990; Riggenbach, 1991; Towell et al., 1996). They were operationalized as follows:

1. Articulation rate (AR): The total number of syllables divided by the total phonation time (excluding pauses) expressed in seconds. Following Riggenbach (1991), the articulation rate was unpruned with all partial words and asides counted. Praat was configured to detect pauses of 0.25 seconds and above.

2. Phonation–time ratio (PTR): The total phonation time divided by the total duration of the speech sample.

3. Pause frequency (PF): The total number of pauses divided by the total duration in seconds of the speech sample. Only pauses of 0.25 seconds and above were used in the calculations.

4. Average pause time: The total duration of all pauses (of 0.25 seconds and above) divided by the number of pauses in a given speech sample.

RESULTS

Perceived Task Difficulty: Quantitative Analysis

All the statistical analyses were conducted using IBM SPSS statistics version 21. The level of significance for this study was set at $p < 0.05$ for all analyses. First, descriptive statistics were computed for participants' overall perceptions of task difficulty as elicited by the five-item Task Difficulty Questionnaire (see Table 1). A Shapiro–Wilk normality test was performed and this showed the questionnaire data were normally distributed. Additionally, measures of skewness and kurtosis were examined and when each was divided by their respective standard error, the result for each of the variables was between +2 and –2. The same tests were carried out for the placement test results, which were also found to be normally distributed ($M = 61.08$; $SD = 12.43$). Based on these results, it was decided that parametric statistical procedures would be used in further analyses.

In order to answer our first research question about differences in the perceptions of task difficulty across tasks, we performed a series of one-way within-subject repeated measures ANOVAs and subsequent paired *t*-test comparisons on the task difficulty perception data (see Table 1). To facilitate comprehension, Task 1 refers to the unrelated picture narration task, Task 2 to the horse-riding story retelling task, and Task 3 to the related 11-frame comic strip. The between-measures ANOVAs yielded a significant effect for two of the five variables, namely: lexical retrieval difficulty (Wilks' Lambda = 0.84, $F[2,38] = 3.74$, $p < 0.05$, partial eta squared = 0.16) and fluency difficulty (Wilks' Lambda = 0.84, $F([2,38] = 3.67$, $p < 0.05$, partial eta squared

= 0.16). Following Cohen (1988), the effect sizes are large, indicating the magnitude of task

effect on the participants' perceived difficulty.

Next, the influence of the tasks was examined by *t*-tests to evaluate the difficulty

differences presented by each task (Bonferroni was adjusted to require $p < 0.015$ for statistical

significance). The *t*-test results determined significant effects for lexical retrieval difficulty

between Task 2 and Task 3 (t[39] = 2.76, $p < 0.015$) and fluency difficulty between Task 2 and

Task 3 (t[39] = 2.69, $p < 0.015$). The results demonstrate a link between task characteristics and

perceived task difficulty. In general, the data show that Task 3 prompted the lowest means across

all the task difficulty perception variables, while Task 2 was associated with the highest means.

Task 1 prompted means in the mid-range.

<INSERT TABLE 1 ABOUT HERE>

Second, descriptive statistics were computed to describe quantitatively the main features

of L2 French utterance fluency measures elicited from the three speech tasks (see Table 2). The

distribution was then verified for extreme outliers, skewed distribution, and unusual patterns for

each task. Next, a Shapiro–Wilk normality test was performed and this showed the utterance

fluency measures were normally distributed. As described previously, the same alternative

method for examining skewness and kurtosis was also applied, and the analysis confirmed the

normality of distribution.

Research Question 2 aimed to gain insights into the interrelationship of task difficulty

perceptions. In order to answer this question, we first correlated participants' judgements of task

difficulty. As can be seen in Table 2, the most significant factors related to overall task difficulty

in all three tasks included perceptions of difficulties in lexical retrieval and in expressing oneself

fluently. Perceived difficulties with grammatical encoding and planning were more strongly

linked to perceptions of overall difficulty in Tasks 2 and 3, and to a lesser degree in Task 1. The analysis also revealed that perceptions of fluency difficulty, lexical retrieval difficulty, and grammar difficulty were highly intercorrelated in all tasks, whereas planning difficulty seemed to be independent of linguistic difficulty in Task 1, which involved the creation of a storyline, unlike the other two tasks where the content was predetermined by the task.

<INSERT TABLE 2 ABOUT HERE>

For Research Question 3, we also compared utterance fluency variables across tasks, and Table 3 presents the repeated measures ANOVAs and paired sample t-test results. The ANOVAs revealed significant differences for two of the four utterance fluency variables assessed, namely, articulation rate (Wilks' Lambda = 0.83, $F[2,38] = 3.83$, $p < 0.05$, multivariate eta squared = 0.17) and average pause time (Wilks' Lambda = 0.82, $F[2,38] = 4.22$, $p < 0.05$, multivariate eta squared = 0.18). The effect sizes were large, showing the strength of the relationship between the utterance fluency measures in relation to task characteristics. Subsequently, significant differences were explored by means of paired sample t-tests (Bonferroni was adjusted to require $p < 0.015$ for statistical significance), which revealed significant effects for the mean fluency measures between tasks. The t-tests revealed significant differences for articulation rate between Task 1 and Task 3 ($t[39] = -2.54$, $p < 0.015$) and Task 2 and Task 3 ($t[39] = -2.32$, $p < 0.015$), and for average pause time between Task 1 and Task 3 ($t[39] = -2.90$, $p < 0.015$).

<INSERT TABLE 3 ABOUT HERE>

Pursuant to Research Question 4, correlations were also computed across tasks to determine whether there was a relationship between participants' perceptions of task difficulty and utterance fluency measures (see Table 4). For Task 1 (unrelated picture narration), the analyses showed moderate correlations between perceptions of fluency

difficulty and articulation rate, phonation–time ratio, pause frequency, and average pause time. For Task 2 (story retelling), the analyses only revealed weak correlations throughout. In contrast to Task 1, correlations with fluency difficulty were very weak in Task 2. For Task 3 (related pictures comic strip), the analyses indicated mostly moderate and weak correlations, with the exception of a relatively strong relationship between perceptions of lexical retrieval difficulty and pause frequency and perceptions of fluency difficulty and average pause time. Moderate correlations were consistently observed between perceptions of planning difficulty and fluency difficulty and articulation rate, phonation–time ratio, pause frequency, and average pause time. Across tasks, perceptions of grammar difficulty had weak relationships with utterance fluency measures. However, one exception was noted, a moderate correlation between perceptions of grammar difficulty and pause frequency in Task 3.

<center><INSERT TABLE 4 ABOUT HERE></center>

Because it is possible that students' level of proficiency acts as a moderating variable in perceptions of task difficulty, we also computed correlations between task difficulty variables and students' placement test scores. The only significant correlation that emerged was in the case of perceptions of grammar difficulty in Task 2 ($r = 0.41$, $p < .05$). This result suggests that, in our study, proficiency did not seem to have a significant impact on how difficult students perceived the tasks to be. The correlations between task difficulty perceptions and utterance fluency data were also rerun using the placement test score as a covariate. This reduced the number of significant relationships, but a large proportion of significant correlations revealed in the analyses presented in Table 4 still remained statistically significant (see Appendix B).

<B>*Perceived Task Difficulty: Qualitative Analysis*

Research Question 1 also sought to address the qualitative differences in learners' perceptions of task difficulty. Table 5 presents the task difficulty features that consistently emerged in the retrospective interviews. As the qualitative data reveal, participants' perceptions of task difficulty are determined by a multitude of dimensions, not all of which are task-specific, as some also relate to individual differences across L2 speakers.

<INSERT TABLE 5 ABOUT HERE>

In total, 135 comments were collected. Although a number of important issues pertaining to perceived difficulty were identified, for the purposes of this study, the qualitative analysis will primarily focus on the comments that can be related to the themes of the task difficulty questionnaire.

The descriptive statistics revealed that the mean for perceptions of lexical retrieval difficulty was the highest for Task 2 (story retelling; 4.18), the lowest for Task 3 (related picture comic strip; 3.60) and in between for Task 1 (unrelated picture narration; 3.88). In general, these results indicated that participants had significantly more difficulty in retrieving words for Task 2 than for Task 1. By comparison, Task 3 was perceived to be the easiest in terms of lexical retrieval. In the retrospective interview comments, certain observed patterns stemming from issues of task structure in relation to lexical retrieval difficulty were noted, some of which were followed by conflicting reports. In the following extracts, participants elaborate on the relation between task structure, degree of freedom or control, and how these affected their perceptions of lexical retrieval difficulty:

I found Task 1 harder than Task 2. I wanted to make an awesome story but I didn't have all the words. Maybe it was harder because there was more freedom. (Participant 11)

Task 2 was harder because I couldn't choose the words myself and there isn't another word for 'stirrup' or 'pond' you can't change it to 'river' or something. (Participant 40)

In contrast, other participants reported some conflicting views on how they believed that task structure and degree of control contributed to overall perceptions of task difficulty:

Task 1 was so much easier because I could pick out my own words and make it go however I wanted. The story retell had a set structure and I know I could go around words if I had to, but in Task 1, it was a lot easier, maybe because I had control. (Participant 13)

Despite the fact that Task 2 had the highest mean for lexical retrieval difficulty, the comments exemplify how individuals vary concerning their perceptions of lexical retrieval difficulties in relation to task structure.

Closely related to the findings on lexical retrieval is the issue of perceptions of fluency difficulty in speech processing. According to the quantitative results, Task 2 had the highest mean (4.20) for perceptions of fluency difficulty, followed by Task 1 (3.80). From the lowest mean (3.58) for Task 3, it is clear that task characteristics played an important role in their perceptions of fluency difficulty. The retrospective interview data support the descriptive statistics. In general, it appears that the specific vocabulary requirements and closed nature of Task 2 contributed to perceptions of task difficulty:

This task was harder. There was no choice in this one. The other tasks were very interpretive but this one is already set so you can only stray so far as to what is written, whereas in Tasks 1 and 3 you can create any story you want. (Participant 15)

However, for other participants, it was the translation aspect and perspective (3rd person singular) of the task that posed some difficulty:

> I found Task 2 harder because I was telling someone else's story, rather than my own, so
>
> I found it more difficult. (Participant 38)

The characteristics of Task 2 explicitly bring issues of task difficulty to the fore, especially because of its closed nature and lack of creative demand placed on the speaker. Nonetheless, the structure, less of a creative aspect and more reporting in nature, was perceived to cause a lower level of difficulty for some students, albeit a smaller majority:

> I felt more comfortable with Task 2 because the storyline was already there and I had
>
> something to go off of. This one was easiest. The task with the unrelated pictures was the
>
> most difficult. (Participant 40)

On the questionnaire, the participants rated Task 2 as the most difficult for fluency, but the qualitative reports showed that some participants considered this task easier than the two more open ended tasks, even though Task 2 demanded that they use more specific vocabulary.

<A>DISCUSSION AND IMPLICATIONS

<B>*Perceived Task Difficulty and its Relationship to Task Characteristics*

In our first two research questions we asked whether there were differences in learners' appraisals of difficulty across tasks and how the L2 speakers perceived task difficulty, both quantitatively and qualitatively. The statistical analyses show that the three tasks varied in perceptions of lexical retrieval difficulty and fluency difficulty. Interestingly, the intercorrelations of the various aspects of task difficulty also reveal that these variables are the most influential in terms of contributing to participants' perceptions of overall difficulty. In the qualitative comments, it is apparent that problems experienced with lexical access are often linked to difficulties with fluent expression. These findings highlight that, in the case of narrative

tasks and within our sample, one of the most important facets of task difficulty to be considered is related to lexical encoding and expressing oneself fluently.

Another important result to highlight concerning task-related variation in perceived difficulty is that the only linguistic aspect of task difficulty participants reported to be different across tasks was related to lexis. It is well known from studies on L2 speech production that lexical breakdowns and difficulties often enter a speaker's conscious awareness (Dörnyei & Kormos, 1998; Kormos, 2006) and trigger problem solving mechanisms, and hence students might perceive and report lexical difficulties with more ease and higher frequency. The findings suggest that students' views of lexical difficulty are salient in task performance, and this underscores the importance of Skehan's (2009) call for more attention to lexis in task-based research. The fact that, in all tasks, lexical retrieval difficulty showed a very high correlation with fluency difficulty, with over 47% (see Table 2) of the variance shared between the two variables, lends support to Lennon's (2000) and Segalowitz's argument (2010) that lexical retrieval is one of the most important factors in L2 proficiency and fluent and stable processing performance.

The findings concerning the lack of significant differences observed for task difficulty items relating to planning and grammar difficulty and overall perceived task difficulty are also worth considering. The quantitative results seem to indicate that when participants are asked to judge difficulty on a 6-point Likert scale and results are averaged—even though the tasks clearly make different planning demands, as indicated by the qualitative comments and the considerable range in the responses (see *SD* values in Table 1)—these factors do not emerge as significantly differentiating between tasks. A similar conclusion can be drawn with regard to overall task difficulty, which, given that it combines judgements on different aspects of task demands, does

not seem to show significant differences for this sample of students. These findings lend support to Bachman's (2002) argument that task difficulty is not a unitary concept and that appraisals of difficulty vary across individuals and communicative contexts. With regard to perceptions of difficulty in relation to grammatical structures, several alternative explanations are possible. On the one hand, it is conceivable that L2 speakers might not be as consciously aware of grammatical demands and their own syntactic encoding difficulties in spoken performance as they are of their lexical problems. On the other hand, the demands of the tasks with regard to syntactic structures might not have been perceptibly different for the learners as all three tasks were of the same genre and required the narration of events.

With regard to specific differences in learners' appraisals across tasks, Task 3 (related pictures comic strip) had the lowest means for perceptions of planning difficulty, lexical retrieval difficulty, grammar difficulty, fluency difficulty, and overall difficulty. In general, the qualitative comments reflect that the structured nature of this task made it less taxing on the speakers because they did not have to plan the story creatively, but rather follow the storyline as depicted in the pictures, for example:

> For me Task 3 was easier than Task 2 in two ways: One, I didn't have to retell it in their words, I could retell it in my own words, so I had control, and, second I liked the pictures because I could look at things in the pictures and this made me feel more fluid when telling the story. (Participant 24)

In comparison, Task 2 (story retelling) had the highest means across all variables. In analyzing the retrospective comments, these results can be attributed to the fact that Task 2 made more cognitive demands on the learners' attentional resources. In this task, the learners first read a short story in English and were subsequently asked to retell it in French. In Task 2, the additional

demands of tense sequencing and specialized vocabulary may have put more pressure on learners' cognitive processing. The fact that the only significant correlation between students' proficiency as measured by their placement test scores and perceptions of task difficulty emerged with regard to grammatical difficulty in this task might lend additional support to this assumption. The retrospective comments also confirm this particular trend. For example, this participant noticed that Task 2 put more demands on accuracy and lexical retrieval:

> In Task 2, I was trying for more accuracy which made it more difficult. It was already in the past tense so I had to think about the forms of the past tense, that I don't think I used completely correctly, and it was someone else's words so I had to translate into first person even though it wasn't first person for me. Tasks 1 and 3 gave me more leeway to express myself how I wanted. (Participant 39)

With regard to the unrelated picture narration task (Task 1), the retrospective comments point to the fact that the increased creative and conceptualization demands of Task 1 made it challenging, as this participant explains:

> I think Task 1 was more difficult because you had to find a way to tie all these unrelated pictures together, whereas in the other tasks, there was already a story there for you. At first I was excited, I get to make it up all on my own, but in the end, it was difficult. (Participant 25)

Relative to Tavakoli's (2009) study in which she examined task difficulty by means of retrospective interviews, certain main themes running through her data analysis on learner performance perceptions reveal similar constructs to those described in this study. If we consider the structure of the four narrative tasks used in Tavakoli's study, we notice some remarkable agreement in the results. Specifically, the ESL learners in her study consistently found the two

tasks that had a fixed structure and a definite sequence to be easiest, while the two tasks that did not have a determined storyline were thought to be more difficult. As reflected in the retrospective interviews of the present study, overall task structure and cognitive and linguistic demand are also consistent with learners' and teachers' perspectives of factors, which impacted task difficulty in Tavakoli's study.

Our findings also complement those of Tavakoli (2009) by showing that, although the fixed structure of the task and predetermined content to be expressed make it easier in the eyes of the learners, tasks in which the accurate and precise rendering of information plays an important role and does not leave opportunities for learners to tailor the message to their own resources, as in our story retelling task, are also perceived to be difficult. Therefore, our results indicate that task difficulty can originate from demands placed on the conceptualization stage of the message (Kormos, 2006; Levelt, 1983, 1993), which is similar to the facet of cognitive complexity in Skehan's (2009) task difficulty model, as well as from linguistic demands in the formulation stage of speech production, as expressed in Skehan's concept of code complexity. It might be worth considering, however, that as our retrospective interview data also indicate, tasks that push learners to express various concepts accurately and precisely might drive linguistic development more efficiently than those in which students can manoeuvre using their current level of knowledge. An additional important issue to note is that the qualitative comments in our study reveal that task characteristics can interact in different ways in oral communicative tasks. Sometimes they might be in conflict (i.e., the need for creativity vs. the opportunity to manoeuvre in terms of content), and on other occasions they can reinforce each other (lexical demands and the need to render given content). Therefore, it is important to consider the interaction of these facets of tasks characteristics when making pedagogical and assessment

decisions.

<B>*Differences in Utterance Fluency Across Tasks*

In our third research question we investigated whether task-related differences can also be observed in the utterance fluency measures of articulation rate, phonation–time ratio, pause frequency, and average pause time. The analyses showed that two utterance fluency variables were sensitive to task-related variation: articulation rate and average pause time. Articulation rate, as its name suggests, is the speed with which one produces an utterance between pauses. Goldman–Eisler (1968; see also De Jong et al., 2012) argues that this variable should be seen as a relatively stable characteristic of individuals that is less prone to variation across tasks and contexts. In our study, however, this speed variable was found to differ across the three tasks. One possible explanation for this result might be that, for L2 learners, articulation rate is not only a measure of the speed with which they articulate an already assembled and linguistically encoded part of their message, but also a reflection of the speed of other speech production procedures that are carried out before articulation. Therefore, articulation rate can be assumed to vary depending on the cognitive and linguistic demands of the task, especially if the whole speech sample is considered for the complete task, unlike in De Jong et al.'s (2012) study, which only assessed the first 30 seconds of learners' performance.

Average pause time, which was one of the measures of breakdown fluency used in our study, was also found to vary across tasks. Interestingly, the results suggest that in the tasks where students demonstrated a higher articulation rate—that is, they spoke faster—they also hesitated for longer when they paused (see Appendix C) for the intercorrelations between articulation rate and average pause time). This suggests that there is a certain trade off between speed and breakdown fluency in our data set, as the sample of L2 learners of French seems to

increase their articulation rate at the expense of taking longer pauses. This phenomenon might mirror the characteristic of pausing patterns in L1 French (Grosjean & Deschamps, 1975). This could serve as an explanation for the differences between our study and those of De Jong et al. (2012), who found that frequency of filled pauses differed between complex and simple tasks in the case of learners of L2 Dutch.

With regard to specific tasks, Task 3 (related pictures comic strip) had the highest articulation rate and longest average pause time. In comparison, for Task 2 (story retelling) the high level of difficulty reported by the learners was reflected in the fluency scores, with articulation rate being the lowest, and average pause time the shortest, in this task. In Task 1, students were also significantly less fluent in terms of articulation rate and average pause time measures than in Task 3.

Task 1, with its high creativity demand, and Task 2, requiring the accurate retelling of a story, both resulted in a decrease in the speed of speech processing. In the case of Task 2, we can theorize that the L2 speakers had more difficulty producing faster speech, possibly due to a lack of sufficient attentional resources imposed by the demands of the task in the grammatical and lexical encoding phases. As for Task 1, the greater requirement of online planning and creativity required to narrate the story reduced the articulation rate and average pause time, whereas the characteristics of Task 3 manifested the opposite effect.

<center><INSERT FIGURE 1 ABOUT HERE></center>

<B>*Relationships Between Perceived Difficulty and Utterance Fluency*

In order to account for variations between learners and tasks, our fourth research question investigated whether the level of perceived difficulty was related to L2 utterance fluency measures across the three tasks. With regard to Task 1, perceived fluency difficulty correlated

significantly with all the utterance fluency measures and overall task difficulty showed a significant negative correlation with pause length. Relative to Task 2 (story retelling), no significant relationship between perceptions of difficulties and utterance fluency was observed. The lack of a relationship between variables of difficulty and measures of utterance fluency might be related to the fact that this task was perceived to be highly difficult for all participants and therefore there might not have been sufficient individual variation in perceptions for significant correlations to emerge. As for Task 3, a number of significant correlations emerged between utterance fluency measures and perceptions of difficulty. A strong correlation between average pause time and fluency difficulty was found and the relationship between pause frequency and lexical retrieval difficulty was also strong. A moderately strong relationship between overall task difficulty and pause frequency and length was also observed. Planning difficulty was related to pause frequency, average pause time, and articulation rate. Those learners who perceived the task to be grammatically more difficult paused more frequently.

The previous analyses clearly indicate that the various speaking tasks show different relationships between perceptions of difficulty and measures of utterance fluency. In the task that required the invention of a storyline (Task 1), both speed and breakdown measures of fluency were significantly related to perceptions of fluency difficulty. In contrast, in Task 3, where the students were scaffolded by the cartoon when conceptualizing the story, only breakdown measures correlated with fluency difficulty. A parallel finding was that, in Task 3, both reported lexical and grammatical difficulties seemed to co-occur with an increase in the number of pauses and shorter pause times. One possible explanation for the different pattern of correlations in the two tasks might be that in tasks that place high demands on the conceptualization phase, speed fluency might also be sensitive to variations in students' perceptions of difficulty. In contrast, in

tasks that do not require extensive content planning, breakdown fluency measures, which reflect difficulties in linguistic encoding, seem to be better indicators of perceived difficulty. Overall, the results suggest that, with the exception of the task perceived to be the most difficult (Task 2), participants' judgements of fluency difficulty seemed to manifest in their actual fluent performance and hence these appraisals can be regarded as relatively accurate. With reference to overall difficulty perceptions, the results are less conclusive, although breakdown fluency measures, especially pause frequency, seem to be related to learners' cognitive appraisals of task difficulty. This finding also supports our previous argument that it is important to collect information about learners' perceptions about specific aspects of task demands because task difficulty encompasses a number of interrelated factors.

<A>LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

The empirical results of this mixed methods investigation have made a further contribution to the study of SLA and fluency by analysing task difficulty through several different lenses. Despite its contributions to the study of task difficulty and L2 fluency research, there are a number of limitations on its scope that need to be considered in future investigations.

First, a major shortcoming resides in the speech tasks themselves. More interactive tasks, rather than structured narratives, might have revealed different patterns in task complexity and utterance fluency conclusions. Thus, one future research direction could be to conduct a similar study but with the use of more open ended, real world, and dyadic speech tasks. This would allow for more freedom on the part of the speakers to demonstrate their skills, elicit enriched data to evaluate task-based variation in fluent speech performance, and facilitate comparisons between studies. Another potential area for future research might be to examine how various

linguistic aspects of task performance, such as lexical variety and sophistication and syntactic complexity, are related to cognitive appraisals of tasks.

The qualitative data also revealed some important findings about the role of communication strategies (CS) and their impact on fluent performance. As a result, research on the use of CS, task difficulty, accuracy, and fluency in classroom and study abroad contexts is warranted. Future research that gathers data from naturalistic exchanges between students and host families or service encounter interlocutors in classroom and SA contexts might also shed light on the development of fluency in different learning contexts. A study focusing on fluency breakdown and the CS that L2 French learners use to cope with performance deficits, perhaps in relation to fluency, might increase our understanding of what it means to be fluent in French and provide a new set of useful correlates that could be used to examine nonnative fluency.

REFERENCES

Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, *19*, 453–476.

Boersma, P., & Weenink, D. (2010). *Praat: Doing phonetics by computer (Version 5.0.25)* [Computer software]. Accessed 9 September 2014 at http://www.praat.org/

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.

Brindley, G., & Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment. *Language Testing*, *19*, 369–394.

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

Coughlan, P., & Duff, P. A. (1994). Same task, different activities: Analysis of a SLA task from an Activity Theory Perspective. In J. P. Lantolf & G. Appel (Eds.), *Vygotskian approaches to second language research* (pp. 173–193). Norwood, NJ: Ablex.

de Bot, K. (1992). A bilingual production model: Levelt's "speaking" model adapted. *Applied Linguistics*, *13*, 1–24.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. (2012). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 121–142). Philadelphia/Amsterdam: John Benjamins.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*, 385–390.

Derwing, T., Rossiter, M., Munro, M., & Thomson, R. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, *54*, 655–679.

Derwing, T., Munro, M., Thomson, R., & Rossiter, M. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*, 533–557.

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

Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 287–313). Ann Arbor, MI: The University of Michigan Press.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215–251.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.

Fillmore, C. (2000). On fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 43–60). Ann Arbor, MI: The University of Michigan Press.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18*, 299–324.

Freed, B. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243–265). Ann Arbor, MI: The University of Michigan Press.

Freed, B., Segalowitz, N., & Dewey, D. (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.

French, L., & Guay, J. D. (2013, June). *The impact of Canada's Explore Program on L2 oral fluency development*. Paper presented at the Canadian Association of Applied Linguistics Annual Conference, Victoria, British Columbia, Canada.

Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching*, *45*, 215– 240.

Goldman–Eisler, F. (1968). *Psycholinguistic experiments in spontaneous speech*. London: Academic Press.

Grosjean, F., & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français: Vitesse de parole et variables composantes, phénomènes d'hésitation. [Comparative analysis of temporal variables in English and French: Speed of speech and compound variables, phenomena of hesitation]. *Phonetica*, *31*, 144–184.

Ishikawa, T. (2011). Examining the influence of intentional reasoning demands on learner perceptions of task difficulty and L2 monologic speech. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 307–330). Philadelphia/Amsterdam: John Benjamins.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, *29*, 24–49.

Jackson, D. O., & Suethanapornkul, S. (2013). The Cognition Hypothesis: A synthesis and meta-analysis of research on second language task complexity. *Language Learning*, *63*, 330–367.

Kormos, J. (2000). The role of attention in monitoring second language speech production. *Language Learning*, *50*, 343–384.

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

Kormos, J. (2011). Speech production and the Cognition Hypothesis. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 39–59). Philadelphia/Amsterdam: John Benjamins.

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

Kormos, J., & Trebits, A. (2012). The role of task complexity, modality and aptitude in narrative task performance. *Language Learning*, *62*, 439–472.

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

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

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *33*, 41–103.

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

Levelt, W. J. M. (1993). Language use in normal speakers and its disorders. In G. Blanken, J.

    Dittmann, H. Grimm, J. C. Marshall, & C. W. Wallesch (Eds.), *Linguistic disorders and*

    pathologies (pp. 1–15). Berlin: De Gruyter.

Mori, J. (2002). Task design, plan, and development of talk-in-interaction: An analysis of a small

    group activity in a Japanese language classroom. *Applied Linguistics*, *23*, 323–347.

Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second language*

    performance assessments. Honolulu, HI: University of Hawai'i Press.

Préfontaine, Y. (2013a). *Fluency in French: A psycholinguistic study of second language speech*

    production and perception. (Unpublished doctoral dissertation). Lancaster University.

    Lancaster, UK.

Préfontaine, Y. (2013b). Perceptions of French fluency in second language speech production.

    Canadian Modern Language Review, *69*, 324–348.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative

    speaker conversations. *Discourse Processes*, *14*, 423–441.

Robinson, P. (1995). Task complexity and second language narrative discourse. *Language*

    Learning, *45*, 99–140.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring

    interactions in a componential framework. *Applied Linguistics*, *21*, 27–57.

Robinson, P. (2003). The Cognition Hypothesis of adult, task-based language learning. *Second*

    Language Studies, *21*, 45–107.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential

    framework for second language task design. *International Review of Applied Linguistics*,

    43, 1–32.

Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In M. P. Garcia–

Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7–26). Clevedon, UK:

Multilingual Matters.

Robinson, P. (2011). Second language task complexity, the Cognition Hypothesis, language

learning and performance. In P. Robinson (Ed.), *Second language task complexity:*

Researching the Cognition Hypothesis of language learning and performance (pp. 3–38).

Philadelphia/Amsterdam: John Benjamins.

Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second

language learning and performance. *International Review of Applied Linguistics in*

Language Teaching, 45, 161–176.

Rubin, H. J., & Rubin, I. (2005). *Qualitative interviewing: The art of hearing data* (2nd ed.).

Thousand Oaks, CA: SAGE.

Révész, A. (2014). Towards a fuller assessment of cognitive models of task-based learning:

Investigating task-generated cognitive demands and processes. *Applied Linguistics*, *35*,

87–92.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. London: Palgrave

Macmillan.

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

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

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy,

fluency and lexis. *Applied Linguistics, 30*, 510–532.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on

foreign language performance. *Language Teaching Research*, *1*, 185–211.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, *49*, 93–120.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183–205). Cambridge: Cambridge University Press.

Tavakoli, P. (2009). Investigation task difficulty: Learners' and teachers' perceptions. *International Journal of Applied Linguistics*, *19*, 1–25.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). Philadelphia/Amsterdam: John Benjamins.

Towell, R. (2002). Relative degrees of fluency: A comparative case study of advanced learners of French. *International Review of Applied Linguistics in Language Teaching*, *40*, 117–150.

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